

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:01 ; Search time 94.8513 Seconds
(without alignments)
6041.216 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MWSEEEEDDGAETQDSE.....KIKGRKSRSHNNKLIQTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archepa:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14333	100.0	2781	4	Q9UG2	Q9UG2 homo sapien
2	2839.5	19.8	2669	5	Q9SVB8	Q9SVB8 drosophila
3	2775.5	19.4	645	11	Q8VDN7	Q8VDN7 mus musculu
4	1557.5	10.9	976	5	Q9W0T0	Q9W0T0 drosophila
5	1368.5	9.5	412	4	Q9H5E0	Q9H5E0 homo sapien
6	871.5	6.1	1711	5	Q45409	Q45409 caenorhabd
7	865.5	6.0	1713	5	Q95208	Q95208 caenorhabd
8	674.5	4.7	1022	5	Q960Y3	Q960Y3 drosophila
9	672.5	4.7	997	5	Q9W0T1	Q9W0T1 drosophila
10	572.5	4.0	2768	5	Q9VC00	Q9VC00 drosophila
11	572	4.0	17352	5	Q95Y02	Q95Y02 procampar
12	560	3.9	5476	5	Q9N317	Q9N317 drosophila
13	560	3.9	5533	5	Q9VPL2	Q9VPL2 drosophila
14	560	3.9	5533	5	Q9U6C3	Q9U6C3 drosophila
15	560	3.9	5554	5	Q9NHN1	Q9NHN1 drosophila
16	560	3.9	5560	5	Q9VPL1	Q9VPL1 drosophila

17	549.5	3.8	6815	5	Q917U4	Q917U4 drosophila
18	549.5	3.8	16215	5	Q9NFS3	Q9NFS3 drosophila
19	534	3.7	3080	5	Q9VRY3	Q9VRY3 drosophila
20	529.5	3.7	3257	5	Q9V736	Q9V736 drosophila
21	521	3.6	2465	5	Q22463	Q22463 caenorhabd
22	508.5	3.5	2897	5	Q9VTD9	Q9VTD9 drosophila
23	494.5	3.5	5327	5	Q76891	Q76891 drosophila
24	488	3.4	2910	10	Q9FND5	Q9FND5 arabidopsis
25	478.5	3.3	2081	10	Q9LH98	Q9LH98 arabidopsis
26	459	3.2	5170	5	Q17490	Q17490 caenorhabd
27	459	3.2	6994	5	Q17343	Q17343 caenorhabd
28	441.5	3.1	4498	5	Q9W273	Q9W273 drosophila
29	437	3.0	5120	13	Q9PU36	Q9PU36 gallus gall
30	434.5	3.0	2977	5	Q9VAP9	Q9VAP9 drosophila
31	431.5	3.0	2402	2	Q9AER7	Q9AER7 staphylococ
32	430.5	3.0	3261	4	Q9Y556	Q9Y556 homo sapien
33	430.5	3.0	3664	4	Q96T58	Q96T58 homo sapien
34	430	3.0	1795	5	Q76894	Q76894 drosophila
35	430	3.0	2586	5	Q9VTK8	Q9VTK8 drosophila
36	429	3.0	3726	5	Q9VPL1	Q9VPL1 drosophila
37	427	3.0	1483	5	Q8T273	Q8T273 dictyostell
38	423	2.9	2285	5	Q9VP17	Q9VP17 drosophila
39	421.5	2.9	5147	4	Q9Y6V0	Q9Y6V0 homo sapien
40	419.5	2.9	1850	11	Q91P55	Q91P55 mus musculu
41	419.5	2.9	2232	5	P91365	P91365 caenorhabd
42	418	2.9	2994	5	Q95ZG5	Q95ZG5 dictyostell
43	415.5	2.9	4025	4	Q9NR13	Q9NR13 homo sapien
44	415.5	2.9	3484	5	P91257	P91257 caenorhabd
45	413.5	2.9	3507	5	Q23587	Q23587 caenorhabd

ALIGNMENTS

RESULT 1

ID Q9UG2; PRELIMINARY; PRT; 2781 AA.
AC Q9UG2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Bromodomain PHD finger transcription factor.
GN BPTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20130111; PubMed=10662542;
RA Jones M.H., Hamana N., Shlman M.;
RT "Identification and characterization BPTF, a novel bromodomain
transcription factor";
RL Genomics 63:35-39(2000).
DR EMBL; AB032251; BAA89208.1; -.
DR HSSP; Q92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT dom.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
SQ SEQUENCE 2781 AA: 311210 MW: 78380C1C4C3CFBE CRC64;
Query Match 100.0%; Score 14333; DB 4; Length 2781;
Best local Similarity 100.0%; Pred. No. 0;

Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MWSSEEEEDGDAEFODSDSEDEDEDEDDDDSDYPEMEDDDDDASCTESSFRSHST 60
Db 1 MWSDEEEEDGDAEFODSDSEDEDEDEDDDDSDYPEMEDDDDDASCTESSFRSHST 60
QY 61 YSSTPGRRKRRVHRPRSPILIEEKDIPLEFPKSSSEDLAMVNEHIMNIAIYEVLNRNGIV 120
Db 61 YSSTPGRRKRRVHRPRSPILIEEKDIPLEFPKSSSEDLAMVNEHIMNIAIYEVLNRNGIV 120
QY 121 LRLSPFFEDPCALYVSOECCTLMAEMHVLLKAVLREEDTSNTTGPADLKOSVNSTLY 180
Db 121 LRLSPFFEDPCALYVSOECCTLMAEMHVLLKAVLREEDTSNTTGPADLKOSVNSTLY 180
QY 181 FIDGMTPPELVRYCESDKRYHHVLPYQEAEDYPYGVEKKIKYLOPLVOPLTNTIARE 240
Db 181 FIDGMTPPELVRYCESDKRYHHVLPYQEAEDYPYGVEKKIKYLOPLVOPLTNTIARE 240
QY 241 ELMSSEGVIOYDHCRCVCHKLGLDLCCETCSAVYHLECVKPRPLEVEPEDEMOCEVCVAHKV 300
Db 241 ELMSSEGVIOYDHCRCVCHKLGLDLCCETCSAVYHLECVKPRPLEVEPEDEMOCEVCVAHKV 300
QY 301 PGYTDCAVLAELQKPKYTRHEPIGYDSRKKYWFNLNRLLIEEDTENENKKIMYSTKVQ 360
Db 301 PGYTDCAVLAELQKPKYTRHEPIGYDSRKKYWFNLNRLLIEEDTENENKKIMYSTKVQ 360
QY 361 LAELIDLDKDYWEABELCKLLEEMREIHRHMDITTEDLTAKAGSNKSFILAAANEELLES 420
Db 361 LAELIDLDKDYWEABELCKLLEEMREIHRHMDITTEDLTAKAGSNKSFILAAANEELLES 420
QY 421 IRAKKGIDINVKSPETERKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDPEPOGK 480
Db 421 IRAKKGIDINVKSPETERKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDPEPOGK 480
QY 481 SEVGDKRSEKSNELSSPAGAGASGSTRILIRLRNPDSKLSQKSOQYAAAAHAEANKL 540
Db 481 SEVGDKRSEKSNELSSPAGAGASGSTRILIRLRNPDSKLSQKSOQYAAAAHAEANKL 540
QY 541 FKSGKEVLVNSOGELSRLLSTKKEVIMKGNINNYFKLGSGKRYVYNOYSTNSFALNKH 600
Db 541 FKSGKEVLVNSOGELSRLLSTKKEVIMKGNINNYFKLGSGKRYVYNOYSTNSFALNKH 600
QY 601 QHREDHDKRRHLAHKFCLPAGEFKWNGSVHSGKVLITSLRLTIQLENNIFSSFLHPN 660
Db 601 QHREDHDKRRHLAHKFCLPAGEFKWNGSVHSGKVLITSLRLTIQLENNIFSSFLHPN 660
QY 661 WASHBRANWIKAYOMCSPPREFALALILECAVAPVWMLPTWREFLGHTRLHRMTSIEREE 720
Db 661 WASHBRANWIKAYOMCSPPREFALALILECAVAPVWMLPTWREFLGHTRLHRMTSIEREE 720
QY 721 KEKVKKKKKKEOEDEETMOQATWVYKTPYVAKHQAOKGEEYRVATGYGWSMISKTHVYRF 780
Db 721 KEKVKKKKKKEOEDEETMOQATWVYKTPYVAKHQAOKGEEYRVATGYGWSMISKTHVYRF 780
QY 781 VPKLPGNTNVNRYKSLLEGTKNNMDENNDSDKRCSPKIKIEPDSEKDEYKGSDAK 840
Db 781 VPKLPGNTNVNRYKSLLEGTKNNMDENNDSDKRCSPKIKIEPDSEKDEYKGSDAK 840
QY 841 GAOONENDISKITEKKDOVKELLDSDSKPCKEEPEVYDDMKTESHVACOSSOVDVY 900
Db 841 GAOONENDISKITEKKDOVKELLDSDSKPCKEEPEVYDDMKTESHVACOSSOVDVY 900
QY 901 NVSEGFHLRTSYKKTSSKSLDGLERRIKOFLLEEKORLEKTKLEGIGIKGTSTNSS 960
Db 901 NVSEGFHLRTSYKKTSSKSLDGLERRIKOFLLEEKORLEKTKLEGIGIKGTSTNSS 960
QY 961 KNLSESPVITKAKEGQOSDSMRQOSPANNDOPEDLIQGCSSSDSSVLRMSDPSHTTNK 1020
Db 961 KNLSESPVITKAKEGQOSDSMRQOSPANNDOPEDLIQGCSSSDSSVLRMSDPSHTTNK 1020
QY 1021 LYPKDRVLDVYSTIRSPETKCPKONSTENDLEEVVSLASGQEPYTSKTYGNDFFIDSK 1080
Db 1021 LYPKDRVLDVYSTIRSPETKCPKONSTENDLEEVVSLASGQEPYTSKTYGNDFFIDSK 1080

```

```

QY 1081 LASADIDGTLICKNKRPILIOESDPTIVSSSKSALHSSVPSKSTNDRDAPTPLSRAMDFEGL 1140
Db 1081 LASADIDGTLICKNKRPILIOESDPTIVSSSKSALHSSVPSKSTNDRDAPTPLSRAMDFEGL 1140
QY 1141 GCDSESNSTLENSSDIVSIOSSSEEDMIVONNESISSEQFRTRODVEVLEPLKCELVSG 1200
Db 1141 GCDSESNSTLENSSDIVSIOSSSEEDMIVONNESISSEQFRTRODVEVLEPLKCELVSG 1200
QY 1201 ESTGNCEDRLPVKGTENGKRRPSQOKKLEERPVAKKGDQIKLNTPTDKKNNEPRESKKG 1260
Db 1201 ESTGNCEDRLPVKGTENGKRRPSQOKKLEERPVAKKGDQIKLNTPTDKKNNEPRESKKG 1260
QY 1261 QRTSTFOINKDKPKIYLGECLEKTESERVVSGANEPVNNINKIIPENDIKSLTVKE 1320
Db 1261 QRTSTFOINKDKPKIYLGECLEKTESERVVSGANEPVNNINKIIPENDIKSLTVKE 1320
QY 1321 SAIRPFLNGDVIMEDFERNSSSETKSHLSSDAEGNYRDSLFTLPSTKESDSTQTTTPS 1380
Db 1321 SAIRPFLNGDVIMEDFERNSSSETKSHLSSDAEGNYRDSLFTLPSTKESDSTQTTTPS 1380
QY 1381 ASCPESNVNQVEDMEIETSEVKKYVSSPTTSEESMLSNDFIDENGLPINKNENNGES 1440
Db 1381 ASCPESNVNQVEDMEIETSEVKKYVSSPTTSEESMLSNDFIDENGLPINKNENNGES 1440
QY 1441 KRRTVITEVTTMTSTVATESEKTVIKVEGDKQVYVSTENCAKSTVTTTTYTKLSTPS 1500
Db 1441 KRRTVITEVTTMTSTVATESEKTVIKVEGDKQVYVSTENCAKSTVTTTTYTKLSTPS 1500
QY 1501 TGSVVDIISVKEQSKTVVTTVTTDSLTGTLVTSMTVSEKYSTRDVKLMEFSREPKT 1560
Db 1501 TGSVVDIISVKEQSKTVVTTVTTDSLTGTLVTSMTVSEKYSTRDVKLMEFSREPKT 1560
QY 1561 RSGTALPSYKRFYTKSKKSIIFVLPNDLKLARKGIRFVFPFNVAKALDIWPPSP 1620
Db 1561 RSGTALPSYKRFYTKSKKSIIFVLPNDLKLARKGIRFVFPFNVAKALDIWPPSP 1620
QY 1621 RPTFGITWRRLQTVKSLAGVSLMLRLMASLRMDMAAVPPGGGSTRTSETETITTT 1680
Db 1621 RPTFGITWRRLQTVKSLAGVSLMLRLMASLRMDMAAVPPGGGSTRTSETETITTT 1680
QY 1681 EIIRKRDVGYGIRFEXCIKKIICPIGVPEPTPKETPPPOKGRSSALARKPRETRKQTG 1740
Db 1681 EIIRKRDVGYGIRFEXCIKKIICPIGVPEPTPKETPPPOKGRSSALARKPRETRKQTG 1740
QY 1741 PVIIETWVAEEELIEMIRFAFERVEKEKAQVBOQAKKRLBOQKPLVITSTTSPST 1800
Db 1741 PVIIETWVAEEELIEMIRFAFERVEKEKAQVBOQAKKRLBOQKPLVITSTTSPST 1800
QY 1801 TSTISPAOKVMVAPISGSVTTGTRMVLJTTKVGSPATVTFQONKNFQTFATWVKOGQSNS 1860
Db 1801 TSTISPAOKVMVAPISGSVTTGTRMVLJTTKVGSPATVTFQONKNFQTFATWVKOGQSNS 1860
QY 1861 GVVVOQOKVLGIIPSSGTSGOFTSFOPRTAVTTIRPNTSSGSGTSSNSQVITGPQIRP 1920
Db 1861 GVVVOQOKVLGIIPSSGTSGOFTSFOPRTAVTTIRPNTSSGSGTSSNSQVITGPQIRP 1920
QY 1921 GMTVIRPILQOSTLGAIIITFPVWVOPGAPQOVMTOIIRQOPVSTVASANTVSSPPGOK 1980
Db 1921 GMTVIRPILQOSTLGAIIITFPVWVOPGAPQOVMTOIIRQOPVSTVASANTVSSPPGOK 1980
QY 1981 SLTSAATSTNSNIQSSASQPPRPOGOVKLITMAOVLTOILQHGNGQGLTVYIOGGQOTTGOL 2040
Db 1981 SLTSAATSTNSNIQSSASQPPRPOGOVKLITMAOVLTOILQHGNGQGLTVYIOGGQOTTGOL 2040
QY 2041 QLIIPQGVTVLPGFGQOLMQAAMPNGTVQRFLETPLATATATTTTASTTTVSTAAGTEGOR 2100
Db 2041 QLIIPQGVTVLPGFGQOLMQAAMPNGTVQRFLETPLATATATTTTASTTTVSTAAGTEGOR 2100
QY 2101 QSKLSPQOMOVHODKTLPRPOSSSVGPAKAPQCPAQSARQPOPTQOSPAQPEVQTPQEV 2160
Db 2101 QSKLSPQOMOVHODKTLPRPOSSSVGPAKAPQCPAQSARQPOPTQOSPAQPEVQTPQEV 2160

```


Qy	2161	QTTTTSVSHSPSAOCPETHNASSKPOVAAASQPSQSNVQGSPPRVQSPSQTIRRSTRSQSL	2220
Db	2161	QTTTTSVSHSPSAOCPETHNASSKPOVAAASQPSQSNVQGSPPRVQSPSQTIRRSTRSQSL	2220
Qy	2221	SPGQSGSVQVTTTSGQPIPIQPHNTSLQIPSGQPSQPSQPOVQSSPQTLSQCTLNOVSVSSPS	2280
Db	2221	SPGQSGSVQVTTTSGQPIPIQPHNTSLQIPSGQPSQPSQPOVQSSPQTLSQCTLNOVSVSSPS	2280
Qy	2281	RPQLQIQPQPOVIAVPOLQOQOVVLSQIQSGVAVQAQAOQSGVFPQIKLQLPQIQOQS	2340
Db	2281	RPQLQIQPQPOVIAVPOLQOQOVVLSQIQSGVAVQAQAOQSGVFPQIKLQLPQIQOQS	2340
Qy	2341	AVQTHQIQNVVTVQAASVQQLQVQQLQFDQOQKKKQOQIEIKRPHTLQASNOSEIIQKQ	2400
Db	2341	AVQTHQIQNVVTVQAASVQQLQVQQLQFDQOQKKKQOQIEIKRPHTLQASNOSEIIQKQ	2400
Qy	2401	VVMKHNNAVIEHLKOKKSMTPAEKEEQNRATVQNOVMKTYTLDKIDKEEQAAKKRKEESV	2460
Db	2401	VVMKHNNAVIEHLKOKKSMTPAEKEEQNRATVQNOVMKTYTLDKIDKEEQAAKKRKEESV	2460
Qy	2461	EOKRSKONATKLSALLFKHKEQURAEILKKRALLDKQIQEVOEELKRDLIKKEKDLQK	2520
Db	2461	EOKRSKONATKLSALLFKHKEQURAEILKKRALLDKQIQEVOEELKRDLIKKEKDLQK	2520
Qy	2521	LAQTAATAACPPVTPLPAPRAPPPSPPPPGVOHTGLSTPPLPVASQKRRKEEKDS	2580
Db	2521	LAQTAATAACPPVTPLPAPRAPPPSPPPPGVOHTGLSTPPLPVASQKRRKEEKDS	2580
Qy	2581	SSSKKKKKMSTTSKKEKKDQTKYCICKIPPYESKPYIGCDRCQMWYGRGVGLQSAE	2640
Db	2581	SSSKKKKKMSTTSKKEKKDQTKYCICKIPPYESKPYIGCDRCQMWYGRGVGLQSAE	2640
Qy	2641	LIDEEVCPQOCOSTEDAMTVLPLTEEDYEGLKRVLSLQAHKMAPFLEPYDPNDADPY	2700
Db	2641	LIDEEVCPQOCOSTEDAMTVLPLTEEDYEGLKRVLSLQAHKMAPFLEPYDPNDADPY	2700
Qy	2701	GVIEKPEMDLATMBEVRQRYEKELEFVADMTKIFDNCRYNPSDSPYQCAEVLESFFV	2760
Db	2701	GVIEKPEMDLATMBEVRQRYEKELEFVADMTKIFDNCRYNPSDSPYQCAEVLESFFV	2760
Qy	2761	QKLKGFKASRSHNNKLOSTAS	2781
Db	2761	QKLKGFKASRSHNNKLOSTAS	2781
RESULT 2			
095VB8			
AC	Q95VB8	PRELIMINARY; PRT; 2669 AA.	
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Nucleosome remodeling factor, large subunit NURF301.		
GN	E(BX) OR NURF301 OR CG7022 OR CG1135.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=21468388; PubMed=11583616;		
RA	Xiao H., Sardaletzpoulos R., Wang H., Hamiche A., Ranaïlo R., Lee K.,		
RT	Fu D., Wu C.,		
RT	"Dual functions of largest nurf subunit nurf301 in nucleosome sliding		
RT	and transcription factor interactions.";		
RL	Mol. Cell 8:531-543(2001).		
DR	EMBL: AF417921; AAL16644.1; -		
DR	FlyBase: FBgn0035118; E(bx).		
DR	InterPro: IPR001487; Bromodomain.		
DR	InterPro: IPR000345; Cyt_heme_bind.		
DR	InterPro: IPR004022; DDT_dom.		
DR	InterPro: IPR000561; EGF-like.		

[illegible]


```

QY 889 VNCQESSQVDVNVYSEGFHL--RTSYKKTKTKSSKLDGLLBRRIKQTFLEEKORLEKIKLE 946
Db 837 -----IDVSKALNAPGRTIYSKAVAKSRDLDLRLKLAEEV--QMAASKI--- 881
QY 947 GGKIGKIGKISTNSKLNSESPIYTKAKEGQSDSMROEOSPANMNDOPEDLIOGQSDS 1006
Db 882 -----PDSM 885
QY 1007 SVLRMSDPSHTTKLYPKDVLDDVSTRSPETCPKQNSIENDIEKVSLASRGCEPTK 1066
Db 886 KPLLVSSQNTMANS-----KOTFLE-----KRLILTE 913
QY 1067 SKTKGNDFIDSKIASADDIGTLICNKKKPLIOESDITVSSSKSALHSSVPKSTNDRD 1126
Db 914 VQAKG----- 918
QY 1127 APLLRAMDEGKIGCDSESNSTLENSDVTJSIODSSEDMIVQNSNESISEQFRTREOD 1186
Db 919 -----GPAWVLELVNSIAKOIQTVRLQFS-----QL 945
QY 1187 VEYLEPLKCELVGSESTGNCEDRLPVKGTGANKGKPSQOKKLEERYPNKCSDOIKLNTT 1246
Db 946 NREAFKVRFC-----YKREC-----NTNSNVSQITQ-----NTCYSPLCQKAR 984
QY 1247 DKRN---NENRESEKKGQRTSTFOINGKDNKPKIYLKGECLKEISESRVSGNVEPKVN 1302
Db 985 AKKELLLLLKKAHTAGNGSKSEYVAAILGAVKKSIL-----EQKILTEKRE--- 1030
QY 1303 NIKKIIPENDIKSLYKESAIRPFIINGDVIEMEPNENSETSKSHLSSSDAEG----- 1356
Db 1031 -----STQVAVVDSEEGKPAESEAPDLIDLQDWHARAHAV 1065
QY 1357 NYRDSL-----ETLPSTK-----ESDSTQTPPSACPSNSNOVEDM-----E 1396
Db 1066 PFDSDSLTECILDQECVYTKIKOEYVNASGCGNTTIPDSOTDSKIDYIESMDVCSNVE 1125
QY 1397 IETSEVKVYTSPTISEESNL-----SNDFTD-----ENGLPINKNEN 1435
Db 1126 IESTESIVITGLNSGNAEDVDMTGMRRKRNQKSKSYIGTKDVLQDTLKDIPLNK--- 1182
QY 1436 VNESKRKIYITEVYMTSVATESKTVIKVEKGDQOTVSSSTENCASKIVTTTTTIVTK 1495
Db 1183 -----QNRREPIT-----ARPVKREC-----VKYERETFEENGNERV----- 1214
QY 1496 LSTPTGSGV---DIISVKEQSKTVYTTVTDLSLTGTGLVTSMTVSKEYSTRDKVKL 1551
Db 1215 YSTSSPRGRVYLLNDAKYLEQA-----VATEDKSTI 1246
QY 1552 MKTSRPRKTRSGTALPSYKREVTKSTKKSIFVLPRNDLKLAKKGITREVPYFNNAKPA 1611
Db 1247 TK--KKSYSR---YPLISFELTHKKRSLVLPRFELLKLARLIGSSSTNGFHAAKN 1300
QY 1612 LDMPPSPRPRTGIMRYRYLQVYKSLAGVSLMLRLMLMSLRWMDMAAKAPRGGSGTRTE 1671
Db 1301 -TIWQVQCSAPLRTCKSYTNSATLSLADLDRILIMSLCMLWMDMAK--PRSTDGKHQV 1358
QY 1672 TSETEITTEIIRKRDVGPYGRFEYCIRKTIICPIGVPETPKETPTPORKGLSSSLRPR 1731
Db 1359 TTDTEIYVTELELKLRRHSGRGKTSYLRRKVVYLPLEMPKRVREV--TSIRSGLR---KRR 1413
QY 1732 RPETPKOTGVIIETWVAEELMELMELRAFAEVEKEKAQAVEQ--AKKLEQOKPTVIA 1790
Db 1414 RAESPQTEPQITEWVDEKLELWEIKFGEQEKARLSAVTRSVASROLE-----A 1466
QY 1791 TSTSTPSSTSTISPAQKMNAP--ISGSYTTGTMVLLTTKVSPPATVITQOKKNFHQTP 1849
Db 1467 SSGNSGNTSGALGVAGRQALPKLSEDVK--EKMEQOKL---ORAVHQQRK----- 1515
QY 1850 ATWVWQOSNGVVOYQOKVLG---IIPSTGTS---OOTTSFO---PRATVYTRP 1898
Db 1516 --LVAIGETIRSTYTPVKGQVIGSRVYIVKNPDGTIRLIQOAVYQVSRGTGANTAAAAAP 1573

```

```

QY 1899 NTSGSGGTTNS-----QVITGPQ-----IRPGMTVIRFP-----LQOSTLGKA 1937
Db 1574 TVGSGTSTQSNPSTSTPHKQVQIIRGPDKVSVGLNPGQOLVMPDGKHLVLTSTSSNS 1633
QY 1938 IIRTPVWVOPGAPQOVWTOIIRGQPVSTAVASADNVSSTPGKSLTSATSTSNIOSSAQ 1997
Db 1634 AGQGNKKVPIKPASTSS-----SPALISSAQTTNPTVPVIKQIAVHKVTKNSATQSIAS- 1688
QY 1998 PPRPQOQVQLTMAQLTQLTQGHGNGLTIVYIQGQQTIGQ-----LQILPGGYV 2049
Db 1689 -----SRRVALPLAQI-----KNKLLLAQOQOQSTSSPAPATSSPQKIVSKVYN 1733
QY 2050 LPPGQOLMAAMPNGVQVGFLEPTLATATATSTATTYSTAA--GTGEORQSKLSP 2106
Db 1734 TSTSGOTLQOVFVQSG-----KLIVGQNNQOGKVIITSTAAQOQGSPIVQOQOLV 1785
QY 2107 QMOVHDKTLPRAQSS--SVGPAKAQ-----PQTLQPSA 2138
Db 1786 SOPIQOS---POQISMVQGNQPTQKVIQOIVNTSNVQOQIVVGGORILISPQOTIWRQ 1841
QY 2139 RPOPTQPOSQAPQEVQOTP-----EVQOTQTVSSHPSBAOPTHAQSSKPOVAAOS 2190
Db 1842 RNVPOSQALOMVQOQIQTOQOQOQHVVQPOQOPVVOSONIYVSSPSAQTKLVKQLVYQ 1901
QY 2191 QPOSNVQOGSPVAVQSPSQT---RIRPSTPQSLSPGQOQVQTTTQSPRIPIQPTSLQ 2245
Db 1902 QSOQTIEKQIQTITTSNETGTQOVLVFNSTLQALQAGKLVQAVTANGQOQVYKPLGNN 1961
QY 2246 -----IPSGO-----POSOPVQOSTQTILSSQOTLNQVSSPSRQLO 2285
Db 1962 AQIVAHIKHOGDGNNAHVTSNSATVAPQANPQSPVKAQALPPOSQOVVQO-----QOQ 2016
QY 2286 IOQPOP-----QVIAVPOLOQOVQV--LSOIQSO-----VVAQIOA 2319
Db 2017 IHQOSPINFESGVYPTIQOQVYLQAVQAPQOQALSYEESLLONPRTGYIKVTAQVQ 2076
QY 2320 QQSG-----VPOQIKQLQLP-----IQIOQSSA 2341
Db 2077 TEHGPRIVLQGLVGNDFTAQOLQVQVQKQLMKAQESNGKLGVPRTKIYLAQVQENA 2136
QY 2342 VQF-----HOLQNVYVQA--ASVQEOLOQRQOLR-----DQQKKKKQO--- 2379
Db 2137 VQSQPPLPVPVHOSAAHQQANNIEIDADLTATYENSTKIDAINNGDOENSKAEVE 2196
QY 2380 -----IEIKREHTILOASNOSEILQOVYMKHNAVY 2409
Db 2197 NSNITTNESPAQSSLSGSEHDEPTNLAGLDSFIDLEKQKQESVYVRYGQIK--AIS 2254
QY 2410 EHLKO-----KKSMTPAERENQRMIV-- 2431
Db 2255 NALKQGNLSPELEKLVCMQKOQENANSTNEMETCSRGSVNEBALTPRSQTDTEWKIRF 2314
QY 2432 -----CNOVMKTIIDKIDKEEKOAKKRRREESVQOKRSQKQATLSALLFR 2478
Db 2315 SLRPNAMTSSQFNRLK-----KNSKNDVAFELGQKOSO-----LER 2355
QY 2479 HKQOLRAEILKRRALLDKDLOIEVQELKRLDKIKKEKDMOLAQATAVAAPCPVPYV 2538
Db 2356 HKELKKNILKRSLSLERNLQSEIHE--DVKTQVRHVRPLSNA----- 2397
QY 2539 PAPPAPPPPPPPVQVQHGLSTPLPV-----ASQRRKKEEK----- 2578
Db 2398 -----SPDEQSENERSG---EPMLDFKRTQEVQNPBRHAGRPKLTTRKKEKLYXCIRT 2446
QY 2579 -----DSSKSKKKKMISTKSETKKDKIKVICICTPVD 2612
Db 2447 PYDQTKFYVQCDICSNWFHEDCVSITEASK--KLSEFLICIDCKRARETOQLQVCSQKQPD 2505
QY 2613 ESKFYIGDRQCNMWHYGRGAILQSEAEILDEYVQCQOSTEDBAMTV--LTPILKEXYEG 2671
Db 2506 ESOFTYICQDKQWPHRGVCGIILQSEAEFLIDEVCECQRRKNDANANKKLTLSNVEEL 2565
QY 2672 KVLRLSLQAHKKAMPLEVPDNDADPYGVVIREPMDLATMEERQVRRYYEKLTEFVADM 2731

```


QY	2509	DLK1KKEDLMQLOAQATVVAACP	EVTT---PVLPAAPAPPPSPGQVHGILLSTPTL	2565			
Db	303	DLK1KKEDLMQLOAQATVVAACP <td>EVTT---PVLPAAPAPPPSPGQVHGILLSTPTL</td> <td>362</td>	EVTT---PVLPAAPAPPPSPGQVHGILLSTPTL	362			
QY	2566	PVASQRRKEEKSDSSSKKKKMI	STSKETKPKDTLYCICKTPYDESK	2615			
Db	363	PASQRRKEEKSDSSSKKKKMI	STSKETKPKDTLYCICKTPYDESK	412			
RESULT 6							
ID	045409	PRELIMINARY; PRT: 1711 AA.					
AC	045409:						
DT	01-JUN-1998	(TREMBLrel. 06, Created)					
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)					
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)					
DE	F26H11.2a	protein.					
GN	F26H11.2A.						
OS	Caenorhabditis elegans.						
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;						
OC	Rhabditidae; Pterodermidae; Caenorhabditis.						
OX	NCBI_taxonomy:6239;						
RN	(1)						
RP	SEQUENCE FROM N.A.						
RA	Barlow K.;						
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.						
RN	(2)						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=99069613; PubMed=9651916;						
RA	none;						
RT	"Genome sequence of the nematode C.elegans: A platform for						
RL	investigating biology";						
RL	Science 282:2012-2018(1998).						
DR	EMBL; Z81513; CAB04197.1; -;						
DR	InterPro; IPR000637; AT_hook.						
DR	InterPro; IPR004022; DDT_dom.						
DR	InterPro; IPR001965; Znf_PHD.						
DR	Pfam; PF02791; DDT: 1.						
DR	Pfam; PF00628; PHD: 1.						
DR	SMART; SM00384; AT_hook: 1.						
DR	SMART; SM00249; PHD: 1.						
SO	SEQUENCE 1711 AA; 196549 MW; AC45F73FD894044E CRC64;						
Query Match 6.1%; Score 871.5; DB 5; Length 1711;							
Best Local Similarity 17.7%; Pred. No. 6.7e-27;							
Matches 420; Conservative 299; Mismatches 618; Indels 1041; Gaps 71.							
QY	4	EEEEEE---DGADETQDSDEDE---EDENEEDDDSDYEEEMEDD---DDDASYCTE	52				
Db	169	KRQEDDIYMDSDSEEBESSDDDEFMLNDQVGEEBELNLTDIKIEKGLDEKNKC--	226				
QY	53	SSFRSHSTYSSSPGRRKRRVHRPSPILFE--KDLPLEFPKSSSEDLMVNPCHIMNVAI	110				
Db	227	-----PWLDEDDPASLPKLELPRESSODIPPTASINDAVEI	261				
QY	111	YEVLNFGCVTLRLSPFREDFPCALVSQEOCTLMAEMHVLKAVLREEDTNTFTGGAD	170				
Db	262	YEILRSYHRTLRITFTFEDFCALISHNNSCTIMAEVHALLRNCLKSDDEQCTHYSTE	321				
QY	171	LKDSVNSTLYFDGMTBEVLRLVRCSDSEKHYHNVLYQEA-----EDPYGYEYENKI	222				
Db	322	TNNSVNIIMIHMDLTLYAEILLQYIEA-----YFPADASYRDAIINVNDYFVGVDANI	374				
QY	223	K-----VLQFLYDQFLTTNTAKRELMSGVALQYQYDHCYVCAK--LGDLLC	265				
Db	375	QRDFSEFFNIKHVFLRVLLFLFMSYRFLYSSSEFKKLNVNNGKQNDENCNCVCGSSGVRVG	434				
QY	266	CETCSAVYHLCEVKKPLLEEVPEDEMOCVECVANH--VPGQTDCAVELOKKKPYIRHPIGY	324				
Db	435	CTQCGAAAHVHEC--SHLRKPFPE-VLVCNICCKNSAVRGVLPDEAVADRP--LRSPPIGR	489				
QY	325	DRSRKKYVFLNRLIIIEEDTENENKKIMYVSTKQVLAELLIDCLDKDYWEALCKLTLEM	384				

Db	490	DRYGRYTWFTYRLRYVQSLDETE----	LYTYSVRLQYLLOKLDRTYYENDLCOTIRL	545
Qy	385	REIHRMDITDLTKNARKSNKSFLAANEELIESIRAK--	GDIDNVKSP-EIEKD	440
Db	546	IDFLEQMALIVENTISERR-----	FAALETMVRKOLIGYDPAEATTPQIYLHARD	594
Qy	441	K-NETENDSKDAEKNREFEFDQSLBKSDDKP-----	DDPEQKSEYDFDKSEKNG	493
Db	595	SMKRMASILBDCAQKGGVQKQEVLRPEGVQSPVKCVQFVEJESILPEMIGF-----		647
Qy	494	ELSPGAGKAGASGTIITRLNPNPSKITOLKQCVAAAHAEANKLKEKEGEVLYVNSQ	553	
Db	648	-----DAKL-----	INTFMSG-----	659
Qy	554	GEISRLSTKREVI-----MKGNIN-----	NYFKLGQGRK--YRVYHNOYSTNSFALNKQ	601
Db	660	-----ATOEEIYEQPVIDISDNEDPASPANILMRGDEGNDQTEMFYIYNYYSRNMENSEFLT	713	
Qy	602	HREDHDKRHLAHKFCILTTPAGEFKW-----	NGSVHSGKVLTITLKLTIOTLENNIPSSFL	657
Db	714	RKKADKKYKMASFE--AQIDNFDWVAVANKRQFYGASLHCKFIMWTLQOYKKNPIDLM	771	
Qy	658	HPNASHRAWMIKAVOMCSKPREFALALILBECANVPVYMLIMREELIGHTRLHMTISE	717	
Db	772	HRKPEPAKGEDELEVSAVDYKKLYVCLIKLDAVAKKTIIFMGMNNGLQIRLEITIDQ	831	
Qy	718	REERKVKYKREK-----	QEEETMQATYVVKYTFPVKHOVMKQGEYRVG--YG	767
Db	832	RENPMKQOURLKIDADALTKLDLDDSFVRVNYMKPKWPTY--	ILRQGETYVNAAGKSMG	890
Qy	768	GWSMISTHYYRFPYKPLGNTNNNYKRSLEGTNNNDMDMSDQSRKRSRPKKIKIEPD	827	
Db	891	GMWAVALAKYVEKWT-----	QVPSPLPLAVYVEELKITSY	926
Qy	828	SEKDEVVGSAAKGADONENDISKITEKKDKDYKELLDSDSKPCKEPEMEYDDDKTES	887	
Db	927	SNRK-----	ARLELLVSKIT-----	942
Qy	888	HVNCQESSQVDVYVNSGCFHLRTSYKKKTKSSLDGLLERIKOTPLEKQRLKIKIEG	947	
Db	943	-----	KKRORSG-----	949
Qy	948	GIKGIGTSTNKSNIJSESPVITAKRGCOSDSEMOEOSPANNDQPEDLIOGCSOSS	1007	
Db	950	-----GASSKKPTELT-----	NGCYSPECRS-----	971
Qy	1008	VLNRSDPSHTTNKLYPKRDVLVDVSIIRSPETKCPKONSIENDIEEKVSDLASRQGEPIKS	1067	
Db	972	-----	NPRKRC-----YSP-----	980
Qy	1068	KTKGNDFIIDSKLASADDIGLICNNKPKLIOESBDTIYSSSKSALHSSVYKSTINDDA	1127	
Db	981	-----	MCNR-----	984
Qy	1128	TPLSRAMDFEGKIGCODESNSSTLENSSDTVIIDSSEEDMIYOUNESISIEQFRTREOV	1187	
Db	985	-----		984
Qy	1188	EVLPEPLKCELVSGESTGNCEDRLPVKGTENGKPKPSOQKLEERPVNKCSDOIKLKTND	1247	
Db	985	-----	GYLAVAKQAHDEKLEEE-----	1001
Qy	1248	KKNNEKNESEKKGQRTSTPQJINKDKMKPKIYLKGBELKJISSRYVSGVNEPKVNNIKI	1307	
Db	1002	-----		1001
Qy	1308	IPENDIKSLVYKESAIRPFINGDVIMEDEFNERNSETKSHLSSSDADAGNYRDSLETLP	1367	
Db	1002	-----		1001
Qy	1368	TKESDSQTOTTPSASCPSBNSVQVQEDMEIETISEYAKKVTSSPTISEBSNLSNDFIDBG	1427	


```
Db 1002 ----- 1001
QY 1428 LPINKNENVGESKRKTITTEVTMTSTVATESKTVIKVEGDKQIVVSSSTENCASVTY 1487
Db 1002 ----- 1009
QY 1488 TTTTYYTKLSTPSTGGSDVIISVKEOSKIVVYTTVDSLTGTTGLTVSMVSKESYSTRD 1547
Db 1010 ----- 1009
QY 1548 KYKLMKFSRPKTKRSGTALPSYRKFEYTKSKTSIFVLPNDDLKIAKKGIREV--PYEN 1605
Db 1010 ----- 1054
QY 1606 YNAKPAIDWPYSPRPTFGITWRKYLOTVKSLAGVSLMLRLMASLRMDMAKYPGG 1665
Db 1055 AGIKSNLLIMPYPAPRPTLDLCWKMOGLNARSLHAVAOLKTIWSSIKFNEFD--PDDT 1111
QY 1666 GSTRTETSET--ETTTEITIKRDVCGPYGIRFEYCIRKIICPI--GVPEYTKERTPQOR 1722
Db 1112 HPDRRVVIDPSSHDERRRRIIRHKEMPEYGOYERYEMEIEIIPLYDEPEEDESMLSRNG 1171
QY 1723 -----LRSSALPKRPET-----PKOT-----GPVIEFWVAEEELE 1754
Db 1172 GSSEFSHRSSAKKRRQREHFLSLKFGNTPKSKNAFSLDNKRKATATREWVGVYTLK 1231
QY 1755 LME-----IRAFAEVEKEKAQAV-----EQAKKRLBOOKPVTATSTTSPTST 1800
Db 1232 VFEIKDWKWIIRAWEAKTAKRLKLEATRKAKADEDERRRIDQOQRSVARIVY--PMHS- 1289
QY 1801 TSTISPAOKWAPRIGSV-----TTGTKWVLTTKVGSPTATVTQON-----KNFHQF 1849
Db 1290 ---LISEERNV--PYLSQOQRRPNNGENRGFLKYNSSSVSQAIGYASTPPGYHQPQ 1345
QY 1850 ATWVGOGSNGVYQVQOKVLGIIPSTGTSQOTFTSFOPRTATVIRPMTSGSGGTTSN 1909
Db 1346 PNIIROAGYNO-----LPKRPITTSPPNFOS--RP--VAIPIPTPOLRAAGADG- 1390
QY 1910 SOVITGPQIIRPG---MTVIRPLQOSTLGKATIRTPVWVQPGAPQVMTQIIRGO---- 1961
Db 1391 --VVRVAMVMPGKSTVNTSTPYQ-----ALNRQOYQDROQOQPAVARLNGHFM 1443
QY 1962 -----PYSTAVSAPNTVYSTPGOKSLTSAN----- 1986
Db 1444 GIMRGGRNPSVOMHRLPQNRALORPGESETTEMRKRVTEALIPDNDGDECPPIPRYD 1503
QY 1987 STSNIOSSASQPRPOGOVKLTMAOLTOLTOGHGNGOGLTVYIGOGGOTTGLOL---- 2042
Db 1504 PTSNFDAQRAQOQHPOSRRPYSTPAQMIKRTTOPGVKH--NVILMKASDGTOKMVLKPGO 1561
QY 2043 IPQGVTVLPGGOQLMOAAMPNGSTVORFLETPPLATTATATSTTTTSTTAAGTGEOROS 2102
Db 1562 FPPGVYI--STGGRVYRQRPYAVQOROLYT-----ATPGTFVYRILPANANGAPRQOQH 1613
QY 2103 KL-----SPQMOYHOK--TLPPAQ-----SSSVGAKAKQPOTAQOSANP 2140
Db 1614 QVRRRVVQASGPRAMEYMDQGTGPPGOQVRYVLGGNSGTPRVNPPKVSRRGPGGILT 1673
QY 2141 QPOTOPSPAPQEVQTOPEVOT---QTVSSHYVSEAQ 2175
Db 1674 MQMVVQOQOQHNPANHMDPDDATGFAVSTTTEQVYDEQO 1711
```

```
RESULT 7
095208
ID 095208 PRELIMINARY: PRT: 1713 AA.
AC 095208:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE F26H11.2b protein.
CN F26H11.2b.
OS Caenorhabditis elegans.
```

```
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT Investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 281515; CAC42289.1; -.
DR InterPro: IPR004022; DDT dom.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF02791; DDT; 1.
DR Pfam: PF00628; PHD; 1.
SQ SEQUENCE 1713 AA; 196824 MW; 178E470017D9AD5A CRC64;
```

Query Match 6.0%; Score 865.5; DB 5; Length 1713;
Best local Similarity 17.8%; Pred No. 1.2e-26;
Matches 423; Conservative 298; Mismatches 613; Indels 1049; Gaps 73;

```
QY 4 EEEEBE---DGDDEETDSEDE---EDMEDEDDSDYPEMEDD---DDDASYCTE 52
Db 169 KRQEDDIYMDDESEDESSDDEFMLNEDQVQVEEELNLTDIRKGLDENNYC-- 226
QY 53 SSFRSHSTYSITPGRRKPRVHRPSILBE--KDIPLEPPSSSELMVNPENHINAVI 110
Db 227 -----PWLDDEPASLEKLEIPSSODIPPTASIMDAVEI 261
QY 111 YEVLNREGTVLRSPRPFDFCALVSQEQCTMAAMHVVLLKAVIREDTSTTPGPAD 170
Db 262 YELRSHYHTRLRITPTTFDFCALISHNNSCTMAEVHMLRLNCLKSDDEQTHSYTE 321
QY 171 LKDSVNSTLYFDGMTWPEVLRVYCESDEKYNHVLRYQEA-----EDYRGPYENKI 222
Db 322 TNSVNMIMHMDLTLYAEILRQYLEA-----YFADASVRDAIVNDYFVGVDANI 374
QY 223 K-----VLQFLVDQFLTNLAREELMSGVQYDDHCYCHK-LQDLIC 265
Db 375 QDSEFPFNKIHVFLVVLFLFMSYRFLYSSEFKLVNNNGKQDNENCRVSGSRVVG 434
QY 266 CETCSAVNHECEYKPRLEVEPEDEMOCCEYQVANK--VPGVDCVAELQKKPPIRHEPIG 324
Db 435 CTQCEAAFYEC--SHLKFPPE-VLVNCTCKKNSAVRGVLPDEAVDRER--LRSQPIER 489
QY 325 DRSRRKYWFLNRLTLEEDTENENEKKIYYSSTKVQLAELIDCLDKDYWEAECLKILEM 384
Db 490 DRYGRYWFIVRLVYQSLDETE---LYYSIVPOLYOLKLDRTYEKDLCDTIRLR 545
QY 385 REEIIHMDITEDLTNKAQSKNSFLAANEELIESIRAKK---GDIDVKSPE-ETEND 440
Db 546 IDEFLEQMALIVEMTSERR-----EAALETMYRQLDGYDEAFATPTQYLIHND 594
QY 441 K--NETENDSKDAEKNREFEEDSLKESDQDKTP-----DDPEQGSSEVGFKSEKSN 493
Db 595 SKKMAASILRDAQKQVQVQVEKLEPPEYGGSPCYQVQVESILPESNIGTF----- 647
QY 494 ELESPEGAKGASGSTRITRLRNPDSKLSQOQVAAAHBAHANKLFKEGKEVLVNSQ 553
Db 648 -----DAKL-----IMFMSG----- 659
QY 554 GEISRLSTKKEYI-----MKGIN-----NYFKLGQBGK---YRVYHNOYSTNSPLNRHQ 601
Db 660 -----ATQEELEVOFVDSIDNFAPASANLIMRGEGDNGDTFMYYNYYSRNEMSESEFLT 713
QY 602 HREDHDKRHLNHLRCLTPAGFEKW---NGSVHGSKVLTITLTLTITOLENNIPSSLT 657
Db 714 RKAADKKKKYMAKSF--AOIDNFDWVAVAKNRFYGDASLHCKFIWMTLOQVITKNIPIDLM 771
```



```

Db      6 NVEIESTEDSVTLGNSGNAEDVDMTPGMRKRKNOKSKSYIGTKDVLDDTLKDIPLNK 65
QY      1433 NEWVNGSKRKVTITEWTMTSTVATESKTVIVKEGDKOTVYSSSTENCAKSTVTTTTT 1492
Db      66 -----QNRFRPIT-----ARPVKREC-----VKYERETENGNERV----- 97
QY      1493 VTKLSTPTGSGV-----DIISVEOSKTVVTTVTDSLTGTGTVTSMTSKEYSTRDK 1548
Db      98 ----YSTSSPGRYVLLNDAKALYEQA-----VKTEDK 126
QY      1549 VKLMKSRPKKTSCTALBPYRKRFVTKSTKSTFVLPNDLKLAKGCIREFPYRNYNA 1608
Db      127 STITK--KPSYSR-----YPLISNFTLHKKRSLLVLPREFELTLARLGSGSSTNGFHAA 180
QY      1609 KPRLDIPYSPRPRTGRTMRVRLQVYKSLAGVSLMLRLMALSLRMDMAKAPPGGSGT 1668
Db      181 KNN-TTWQYCCSRPLRTCTKSTYTSNATSLSSIALQRLITMSCLRMDMDMAK--PPSIDK 238
QY      1669 RTESETEITTTTELIRKRDGVPYGIREFECIRKLIQPIGVETPKETPTPORKGRSSAL 1728
Db      239 HOYTDEIYVTLLEKLKRNHSGRYGKESYLRKRVVIPLEMPKTVREV-TSIRGLR----- 293
QY      1729 RPKRPETPKOTGVYIETWVAEELEMLERAERKEKKAQAVEQO-AKKRLDQOKPT 1787
Db      294 KRRRAESPOETPEOITEEWDEDEKLELMEIKFGEKEKARLSAVTBSVASROLE----- 348
QY      1788 VITASTSPTSTSTTSPAKVWVAP-IGSVTTGTGKVLTKRVGSPATVTPQOKNPFH 1846
Db      349 --ASGSGNSTINGALGVAGRVQALPKLEEDK--EKMDQOLK--QAAVHQORK--- 398
QY      1847 QTEATWVKOGOSNGVVOYQKVLG-----IIPSTGTS-----QOFTSFQ-----PRATVT 1895
Db      399 -----LVATGEITRSVTPVKGVIYSRVLVKNPDGTRIIQOAVNQVSTGANGTAA 453
QY      1896 IRRPTSSSGGTS-----QVITGPQ-----IRGMVIRP-----LQOSTL 1934
Db      454 ASPVGGSTGTSNPTSTPHKVOIIRGPGKVSVRGLNPGQOLVQMDKSLHVTITTS 513
QY      1935 GKAIIRPVWVQGAPOQVMTQIIRGQVSTAVSAPRTVSTFGOKSLTATSTNIQSS 1994
Db      514 SNSAGQGNKKKVPRIKASTSS-----SPAISAQTTNPVPIVQLQIAVHTNNSAQSI 569
QY      1995 ASQPPRPOGQVKLTMAQTLQFGHGNQGLTVIIOGGQOTTGQ-----LQLIPOG 2046
Db      570 AS-----SSRVALPLAQI-----KNKLLAQOQOQSTSSSPATSSSPVQKITVSK 613
QY      2047 VTVLPGRGQOLQMAAMNGTVQRLFTPLATATTTATTTVSTTAA---GTGGEQROSK 2103
Db      614 VVNTSTSGQTLQOVFVQSGS-----KLIVGONMAQGVITISTSAQOQGTSPVQOQO 665
QY      2104 LSPQMVQNHOKTLPPAOSSS-----VGPARK--AQPTAOPASARPOPTQPOSPAPPEVQ 2155
Db      666 LVGQSPRIQGS-----PQGISMTQOQIIVGGGRILILSPQGITVTRQNVQOSQALQVQOQIQ 721
QY      2156 TOP-----EVQOTVTSVSHVPESEAPRIHAOSSKQOVAAGOSPOSNQGSFVRYQSP 2207
Db      722 TQOQOQOQHNVVQPOQOQVVOQSNQIVQSSPSAQTKLVKQVLVVOQOQSOQETIEKQITPTTDS 781
QY      2208 SQT-----RIIPSTPQSLPSGGQOSQOVGTTTSOPRILPHTSLQIPSPGQGOQOSPOVQOST 2262
Db      782 NETGTQOVLVNSTLQAOGLAQKIQVATVNGQOQVYAP-----LNNQAOIYAH 831
QY      2263 QTLSSGQTLNOVSVSPSRPOLQIQ--QPOPOVIAVPOLQOQOVV--ISOIOSOVAVQIOAQ 2320
Db      832 KHQDGNALHVTSSATAYVAPQANTSPVKQOALPQSPQOQVYVQOQOQHQOQSTPNEESG 891
QY      2321 QSGVPOQIKL-----QDPIQIQOSSAVQTHQION-----VTVCAASVQ----- 2359
Db      892 VTPITQOPLVLOAQVQAPQ--QOALSYEESLQONQPGTVIKVTAQVLOTEHGPRIVIOG 950
QY      2360 -----EOLQRYOQLRQOQKQKQOQ 2378
Db      951 LVGNDFTAQOQLOQVOTQOVKQOQLMKAQE 977

```

```

RESULT 9
Q9W0T1 PRELIMINARY; PRT: 997 AA.
AC Q9W0T1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CG7022 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
DR EMBL: AE003467; AAF47362.1;
DR Flybase: FBgn0035118; E(bx).
SQ SEQUENCE 997 AA, 109285 MW, 8106E06C8FE6A8C7 CMC64;

```

Query Match 4.7%; Score 672.5; DB 5; Length 997;
 Best local Similarity 26.1%; Pred. No. 3, le-19;
 Matches 283; Conservative 132; Mismatches 343; Indels 327; Gaps 45;

```

QY 1394 DMEIETSEVKKVITSSPTTSEESNL-----SNDFID---ENGPIPK 1432
Db      6 NVEIESTEDSVTLGNSGNAEDVDMTPGMRKRKNOKSKSYIGTKDVLDDTLKDIPLNK 65
QY 1433 NEWVNGSKRKVTITEWTMTSTVATESKTVIVKEGDKOTVYSSSTENCAKSTVTTTTT 1492
Db      66 -----QNRFRPIT-----ARPVKREC-----VKYERETENGNERV----- 97

```



```

QY 195 CESDKEXHNVLPYOEADYRPGVFNKIKVLOFLVDOLFTTN----- 236
Db 164 NKGS-----CCPVRSYSDH-----ENELD-----FMDOSTTTTVAPTTGFIILASTMTP 209
QY 237 -----IAREELMSEGV-IQYDHCRCVCHL-GDLIC-----CE-----TCSAYVHL 275
Db 210 PTTTDCIHDEIFADGSLGKNACEHCYCMRDIVCAVOCECPMMAANGKSCRAMPAA 269
QY 276 ECKAPLEVEPDEMOCEVCAHKVPGVTCV-----AELCKNKPVIRHEPIGYDRSR 329
Db 270 EGECOBPNVYCEDSSSTEVEITTPESATSVPAKGIHAIPKEDVDLOHIDDNDKKE 329
QY 330 KYWFLNRLIIEEDTENENKIMWYSTKVOLAELIDCLDKDYW---EAEIOLKILEMRE 386
Db 330 TATIPSAELSGEIEEBEEDK---KATVVAPOVTD--EKDPSFPESSTAGIIPSDSR- 383
QY 387 EIHHRMDITDILNKAGSKSFLAANEILISIRAKK---GDDIDNVK----- 432
Db 384 -----IDLPSSTSEESKESSTEAAEEDIVKIVTTPPEGSGEEDVPRKSOIPEKEIT 434
QY 433 -----SPEE-----TEKDKNET-----ENDSKD 450
Db 435 EDELIKVTSAPAKASPEEVEVYATTSAPTEEDVKPTTACTISEEBEGKPTPAEBGSGE 494
QY 451 AEKN-----REEFEDQ-----SLEKSD-----DKTPDDPEOGKSEV 483
Db 495 EKKDVKTAAPEEDEDPAKPTSAVASEDEKQPKPSESGDEELDLKPTTATAGATSA 554
QY 484 GDRKSEKSNELSESP-----GAGKASGSTRITLIRNPDLSOLKSOOVA 531
Db 555 SESESEODECKSTEAPSYVDLIEPAPKTESSEBASGEDEVAKETTPAGASIASAGEEIV 614
QY 532 AAHEANKLFEKEEVLV-----VNSOGEISRLST-----KKEVIMK 568
Db 615 KGTTPAGEPSEDEDEIVKGTTPAESSESEBELIKVTTTPAGEPVSAGEELAKETTPA 674
QY 569 GNINNTFKIGQEGKRYVYNHQYSTNSFALNKHQREDHDKRRHLAKFCLTPAGEFKWNG 628
Db 675 GEPS-----IAGEEIVKV-----TTPAGESSIAGEEIVK-----VTPPAGESSSEG 717
QY 629 SVHGSVLT-----ISTLRITITQLENNIPPSFLIPNMAHR 665
Db 718 EEBITIKVTTTPAGSSSEBDEIVEIKESTPAGEPISSEBEDIKATTSAPASDIEGVKPEPT 777
QY 666 ANNIKA--VOMCSKPREFALALAILLECAVAPVYMLPIWREFLGHTLRHMTSIEREKEK 723
Db 778 ATEVPAEVEDFAKP-----TTPIAAEEBPIAGTPIPIPDGIS-----GEEEIVK 822
QY 724 VKKEKKQEBEETMOQATWYKTFPVYKHQVYKOKGEEYRVYTGWSWISKTHVRFVPR 783
Db 823 GTTPOLLEBOPELISEBTEV---PVAED-----D 848
QY 784 LPGNTVNVN-YRKSLEGTKNMDEMNMDESDKRCSRSPKKIKIEPDEKND----- 831
Db 849 LSSSTASAIASSTEGVODASTTTSAPARAGDKDEAATVYPTAODKDEVEVEDATDL 908
QY 832 ---EVKGSAAKAGADONEMDISKITEKKDQVYKELIDSDSKPCKEPEMVEDDMK---- 884
Db 909 PVEDVYOSTTAKATTTEQPKREESSTEAADAIEIVTTSRADQVEVPEADBDHDEED 968
QY 885 -----TESHVNCQESSQVDVNVVSGFHL-RTSYKKTKTKSSKLDEL 924
Db 969 VQTATDLPKISDIPPVVTEATTGQPETSDETATDKPPSYLDPPVSOEVPSTAKVDNR 1028
QY 925 LEHRIRKQFL-----EKKORLEKIKLEGIGK-----IGKSTSSKNLSSEPVITTA 972
Db 1029 NDETETKPTLPPSGEDOSSEPTPAMDLPAIGPEGDLVEGKTYANNTTVPATAPDVASC 1088
QY 973 KEGCOSDKROGESPANNDQPE-----DLIOGCSOS---DSSVLRMDSHTTKLVP 1023
Db 1089 K--CISISIVACQOMEKLEPENLEKCTVADDLIDGCCPTIYICDESTISAEKDESTAK--P 1144
QY 1024 KDRVLDVDSIRSPTKCPKONSIENDIEEK-----VSDLASRGQEPRTSKTKT 1070

```

```

Db 1145 DNKIDDEVSEISTE-ELPKVIMPTGTTEBPLSHVAKPDERIQVTVTSVPAQDFDSTTAKVD 1203
QY 1071 GNDFFIDSKLASADDIGTLICKNKPPLIQEESDTJVSSKSAHSSVPKSTNDRDATPL 1130
Db 1204 KKP--IDIS-----ADKKRPIGSEED-----SKPIDESEEDKKPV 1237
QY 1131 SRAMDEGKIGCDSSESNSTLENSDPTVSIODSSEEDMIV-----QNSGESISEQPRR 1183
Db 1238 E-----ESAEKPKVDESEKRPLETVIPASEIEKESKPEDEKKT 1278
QY 1184 EOD-----VEVLEPICKELVSGESTGNCEDLVPYKGEANAK---KPSOOKKLEERPVNKC 1236
Db 1279 EADFAAPTEQPEATPAQIADTAKEVEDKLATTSAPVSEEDBLKRADEKKRE----- 1332
QY 1237 SDOI---KLKNTDKNKNENR---ESEKQORTS--FOJNGDKNKPVIYKGECLKET 1287
Db 1333 TAOIPPAEIPASTDEPESSTELPTVDLKKPEEDSSTKGEVAPESDKVPEVPTASASTENEI 1392
QY 1288 SESRVVSGNVEPVNNINKLIIP--ENDIKSLTYKE-----SATRPFINGDYI----- 1332
Db 1393 EESDKFTTVAAPKISASDETEPTAEEDLVPAITEPIESEFEVSTKPAVOGPPPLPLA 1452
QY 1333 -MEDFNERNSSEKSHLLSSDAEGNVRDLETLPTKESDSTOTTP--SASCPESSNVN 1390
Db 1453 QPEKKPVADTETSTEADISTPESAVEKEKASGETSESDNEIDGASSTVPVVSADDEKTPS 1512
QY 1391 QVEDMEIETSEYAKVTSPTT--SEESN---ISNDFIDENGLPI-----KNENVNGE 1439
Db 1513 TEKTVE---ADKFTTVAPLAGDEESNLPKLPDIFEEBA--PVAVTTAAPSDD--GE 1565
QY 1440 SK-----RKTVTEVTMTSTVATESEKTVYKVEKGDKOVVSTENCAKSTV 1487
Db 1566 QKPEVEEKPRIEQCPRIEDETSTPS-----SENTEPESDRATTIAPSKEE--PSEPS 1618
QY 1488 TTTTIVTKLSTPSTGSGVDIISVKEOSKTVVTTVTDSLTTTGGTLVYTSMTVSKYSTND 1547
Db 1619 TGAPTKDEPAEPSTDAPEDES--KETPESEVPTTVA--GEKIPSTSIPTDEEPTAT 1673
QY 1548 KYKLMKFSRPPKTRGSTALPSTYKFTKSTKSIPLPBDLKLKLARKGIEVYFENN 1607
Db 1674 SAVAPKPDDEVEKETESTETPTDAPASSEEDENS---STDQIPS-----EVP---E 1717
QY 1608 AKPALDIWYPSRPFTGTWRYRLQTVKSLAGVSLMLRLMASLRWMDMAKVPFG-- 1665
Db 1718 KKP-----BTPAQTPREGDI 1732
QY 1666 -GSTRETSETETTTETIIRKRDVGYGIRFEYCIRKIICPIGVETPRETPPQOK--GL 1723
Db 1733 VCATAPPTSDEVPYQURLPE-----EVLAEIIPQSTETGI 1768
QY 1724 RSSALPKRPPEPKQGPVILIEVWAEELMEIAPAEVREK--EKOAOVA-----Q 1776
Db 1769 KOODETTAAPSIDRK-----EPYVEIDEATVYAPISEKDEKPTEEKPVEQKPTGEE 1822
QY 1777 AKRLEQOKPVYLAISTSPSTSTSTISPAQVWVAPISGVTGKMLTTKVGSPAT 1836
Db 1823 PSBEKEKEPIQDVGSTEBPVGSTEASEASTES-----SEEVKSTBEVAKREED--- 1873
QY 1837 VTFQKNKFNHQTPAIWVKQGSNS--GVVOYQOKVLGIIPSSTG---TSQOTFTSFQPTA 1892
Db 1874 -----KOPSSTAQAPVETIPEISTELPAQOGDKRPTSEAPVSDSDETS 1916
QY 1893 TYTIR-PNTSSGSGGTTSSNSQVITGPQIRGCMVYIRPLQOUSTLGKAIITPVMVPGAPQ 1951
Db 1917 PSEDEKIPSVSG-----EEVEGPEV-----TTASPOAAEDELKTPPESEBSSTD 1960
QY 1952 QV-MQOIIIGOPVSTAVSAPNVTYSSPPGOKSLTSATST-----SNQSSASQPPRPOQO 2005
Db 1961 KYPETTYQKPEBETRADEPESVTOVSVAITSTSAVAGGDIKEDQATTAAP--EEBEE 2018
QY 2006 VKI/TMAQLTQLOUGHGNOGLTVVIOGQOTTGOLQLOIPQGVTVLP--GPGQOLMQ--AAM 2062

```



```

Db 2019 IKPTIAPAEIQQ-----PSEKEPVDE-OEVESGKATPAEEDGCPIDEIAPA 2065
QY 2063 PNGIVQREFLEPLATTATTAATTTTAAAGTGEOBOSKLSPOWQVODK-----TL 2116
Db 2066 TSGPIDE-----ASTAAPTKEESTIVASAA-----SP-AVHDEIDKDVTTTQ 2106
QY 2117 PPAOSSSVGPAKAQOTA-----QPSAR-POPOPOPOSPAPREVOTOREVOT--QTTV 2166
Db 2107 PVADKEVNAAPODETKTSIDVSTDSPTADDEKODTEAPVAPTTVSSPTAASADSSSTP 2166
QY 2167 SSHVPS-----EAQPTN-----AOSKPOVA-AQOSPOSNVGOSSPVRY--OSPQTRIRPS- 2215
Db 2167 TVEVSPSVEIDTRKPMDDIMSQTIIAPHTADGASTSEDEDAVTVSPDAEKTPTVSPAP 2226
QY 2216 -----TPSOLSPGOOSQOVOTTTSQ-----PIPIOPHTSLQIPSGOGPOPOVOSTOT 2264
Db 2227 QOSDMPSESEAPQADAEIPATATPIDDNKIPATVAPOTDGVPATAPRLDEDKIQTTAP 2286
QY 2265 L-----SSGQTLNOVSVPSPSPQLOIQOPQOVIAPQLOOQOVLSQIOSQVVAQIQ 2318
Db 2287 LDEEKIPSTAAFLDDEKIPAPVSPVFDVEPSESEKPAVSEYDGE-----ESTERPVDVE 2341
QY 2319 AAGSGVPOQIKILOPQIQIOSSAVQTHQIONVTVQA-----SVQEOIQORVQLDDQQ 2373
Db 2342 TSTDEPTSAKLKPTPSAPATPSESAPATAEIVPETAPELEKEVEKATEQOPELEKTP 2401
QY 2374 KKKOOIEIKREHTLQASNOSEI-----IQOVVAKHNAVLEHLKOKKSMTP 2420
Db 2402 EKATQEPLEKTPREKATQOPELEKTPREKATQOPELEKEVDKATEQOESVDEKTTPEP 2461
QY 2421 AER-----EENORMIVCNOVMKYLIDKIDKEKQAKRRKRESEVQKRSQONATKLSA 2474
Db 2462 VVKPSLIDSTEED-----ESVSESEESADKKDKNKETEDTDKHEEPEVA 2508
QY 2475 L-----LFKKHQLRAELIKKRALLDKDLQIEVQELKRLKIKKKEK 2516
Db 2509 VVSEIPQOSEAVPPTGHPLEPH--LASSTTTPPAVDR--VGEE-----DEEN 2553
QY 2517 DLMQLAQAFAVAPCPPTV-----PVLAPAPPSPPPPPGVOHGTLSPTLPVASOKR 2572
Db 2554 TVVKLSSSTTTSTTESPVASSTTTVASOQOOQPIPPRYG--HA-----PEDEY 2603
QY 2573 KKEEKDSSSKKKKKMISTSTSEKTKKDKTKLYCICKTPYDESKYIGCDRCOMWYHGRCV 2632
Db 2604 DEEEVFGPTCYAGKLYVSAQOIPRDPCDFCFC-----FRSDII 2644
QY 2633 GILQS-----EAEILDEIVYCPOCQ-----STEDAMFVLTP 2662
Db 2645 CLOQSCPPPIACGHEPISGFCPCRYECPVSMAAVLNITTTSTTTLPP 2695

RESULT 11
095YM2 PRELIMINARY: PRT: 17352 AA.
AC 095YM2:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE I-connectin.
GN I-CON.
OS Procambarus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidae;
OC Astacidae; Cambaridae; Procambarus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomeres of crayfish claw muscle.";
RL EMBO J. 20:4826-4835(2001).

```

```

DR EMBL: AB055861; BAB64297.1; -.
DR InterPro: IPR000577; FGGY_kin.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 48.
DR Pfam: PF00018; SH3; 1.
DR PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_3.
DR PROSITE: PS50002; SH3; 1.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 4.0%; Score 572; DB 5; Length 17352;
Best local similarity 17.7%; Pred. No. 8, 9e-14;
Matches 576; Conservative 570; Mismatches 1232; Indels 874; Gaps 133;

QY 3 SEEEEDDGAETQOSEDEDEDEDEDDDDSDSYPREMEDDDDDAATCT-----ESSFR 56
Db 12294 STEISDKKPKSEKATSIISQESISVQESIVKDAPOSIKDAKPTQATSIISPHSLTI 12353
QY 57 SHSTVSTP---GRRKPRVHRPRSPILDE---KDIPLEPKSSEDLMPNEHIMNY 108
Db 12354 QEVSVKETPTDISDKRPRTEKATSIISQESISVQESVKEAGVETAKPTQATSI 12413
QY 109 AIYELRNFGVLRISPFREDFCALVSQEOCT-LMAEMHVLLKAVLREEDTSMFTFG 167
Db 12414 SPHEISIQEISVKEAP---TDICDKPKSEKATSIISQESIAVQEVSVKE-----A 12463
QY 168 PADLKSVNSTLYFDIGMTWPEVLRAYCESDKETYNHVLPRQEKEDVPYGVENKIVLQF 227
Db 12464 PGSIEEAKPT-----EYAKSSISPHSLTVQEV----- 12492
QY 228 LVQDFLTNMLAREBELMSEGVIOYDHCRCVCHKLGLDLCETGSAVYHLCVKKP--LEEV 285
Db 12493 -----LYKESAEISDKPOTGATSIISQESIAVQEVSVKEKPGVEDV 12538
QY 286 -PEDEMOCEVCAHKKVPGVTDV---AEIOKNKPYIRHEPIGYDRSRKKVFLNR-- 337
Db 12539 KPTQEGATSIISHQSLTVQEVTVKGDASAEISDKP-----KTEKATSIISQES 12588
QY 338 LIIEPTENDNEKKIVYISTKVQVLAELIDCLDKDYEAELCKILEMREIEIRHMDITED 397
Db 12589 ISVQELTVKAEPSVKEDEKLKPEQA-----TSVISPLESITVQEVSVKESSD 12636
QY 398 LTKKANGSNKSFLEAANEELT---SIRAKGDIDNVKSPER-----TEK 439
Db 12637 ISRRKPTKATSIISQESISVQESIVKAEAPCTVADVPKPEQATSIISQESLTVQEV 12696
QY 440 DKNETENDSKDAEKRE---FEDQSLD---KDSDDKTPDDP--EQGKS----- 481
Db 12697 TVKEIPTDISDKPKSEKATSIISQESISVHEVSVKDAPGSMKDAKPTQATSVISPH 12756
QY 482 -----EVGDF--KSEKSNGLS-----ESPAGKAGASTRI 511
Db 12757 ESTLVQEVSVREVPTEISDKPKSEKATSIISQESISVQESIVKAPGSMKDAKQTEQ 12816
QY 512 ITPRLRPDSKLSQKQVAAAHNEANKLFKEGKEVLVYVNSQGEIRLSTKKEVIK---- 568
Db 12817 ATSVISPHSLTVQEVSVKEAPTELSDRKPKKEKATSIISQESIS---VQEVSVKEAP 12872
QY 569 GNINNYFKLQGEKKYRV--HNQYSTNSPALNKHQREDHDKRHLAHKFCLLPAGEFWM 626
Db 12873 GSVKD-LKLTQEGATSVISPHSLTVQEVSV--KEATTEISDKK-----PKTE-- 12917
QY 627 NSVHSGKVL---TISTLRLLT---QLENNIPSFILPKNAS--HRANMIKAVDMCSK 677
Db 12918 ---QATSIISQESISVQESIVKAGPGLKDEKPKSEQATSIISHQSLTVQEVTVKEA 12973
QY 678 PREFALALALICQAVPVVMLPIWREFLGHTRLRHRTSLIEREKE-----KVKK----- 726
Db 12974 PADIS-----DKP-----KTEKATSIISQESISVQESIVKKEAPGSM 13011
QY 727 KKKQDEEE-----TMOQATWVRYTPPVKHQVWKQGEVRYVYGGSWISKT 775

```



```

Db 13012 KDAKPKTQATSVISPHESLTVQEVVS-VK-----EVPTEISDKKPKSEKATSIILSEQESISV 13067
QY 776 HYREFVPLPGNTNVNKRSLGTR-----NMMDNMDESKRKRKSRPKIKIPEPD 827
Db 13068 HEVVS-VKSPSGMKDAKKTQATSVISPHESLTVQEVVSKEPTEISDKKPKSEKATSI 13126
QY 828 -SEKDEVKGS-D-AKGAQONEMDISKITEKKQDVKEILD--SDSDKPKCKEPEMEVDDM- 883
Db 13127 LSEQESISIOEVSVEKPEMSKDLKPKTEQATSVISELISLTVQEVVSKEPTEISDKKPK 13186
QY 884 KTE---SHVNOESQOVVAVVSEGFHLRTSYKKTK--SKLDGLERRIKQFTLEEKQ 938
Db 13187 KTEKATSIILSEQESISVOEISVKEAPGSVEEVKPKTEQATSVISPHMSLTVQEVVKE-- 13244
QY 939 RLEKIKLEGIGIKIGIKTSTNSSKILSESPVITAKEGQOSQMSQMOEOSPANNNOQPDLI 998
Db 13245 -----VPADISDVKPKSEKATSIILSE-----QESISVOEISVKEAPGSMD----- 13285
QY 999 QGCSQSDSVLRMSDPSSHHTNKLYPKRQVLDDVSIRESPEKCPKQNSIENDIEEKVSDLA 1058
Db 13286 -----AKPKTEQATSVISPHESLTVQEVVSHEVPTEISDKK 13321
QY 1059 SRGQEPKSKTKGNDPFIQDOKLASADDIGLICKNNKPLIOESDITVSSKS-----A 1113
Db 13322 PKSEKATSIILSEQESISVOEISVKEAPE---SKMDAKPK--TEQATSVISPHESLTVQEV 13376
QY 1114 LHSSVPKSTNDRDATPLSRAMDFEKLGCDSNSLTENSSTLENSSDVSIQDSSEEDMIQVNSN 1173
Db 13377 SVKEVPEISDKP-----KTEKATSIILSEQESISVOEVS-----VKEAP 13416
QY 1174 ESISEQRTHEQVVEVLEPLKCELVSESGNCEBRLPVKGTANGKPKSQOKL--E 1229
Db 13417 GSVADLKPKEQATSVISP-----HESLTVQEVSVKEAPTEISDKKPKTEQATSVILE 13469
QY 1230 ERPVNCSDQIKLNTDKNNENRESEK-----KGQRTSTFOJNKG-----DNKPK 1276
Db 13470 QESIS--IOEISVKEAPGSLKDEKPKSEQATSIILSEQSLTVQEVVYKEAPADISDKP 13527
QY 1277 IYLKGECLKF-----ISSESVVSGNVEKRVNNIKIPENDIKSLTVKESAIRP 1326
Db 13528 TEKATSIILSEQESISVOEISVKEAPGSMKDAKPKTEQATSVISPH--SLTVQEVSVR-- 13583
QY 1327 INGVIMEDENRNSSEFTKSHLSSSDAEQNYDSLETLPTSESTQJ--TTPASCD 1384
Db 13584 ---EVPEISDKKPKSEKATSIILSEQESISVOEISVKEAPGSMKDAKPKQATSVISPH 13640
QY 1385 ESSNSVQOVEDMEIETS---EVAKVTSSPTSEESNLSDNDFIDENGLPINKNENVNGES 1440
Db 13641 ESILTVQEVSVKEAPTELSDEPREKEKATSIILSEQES--ISVQEVSVKEAPEMSKEAPTE 13699
QY 1441 KRTVITEVITMTSTVATEKTYIAKEKGDQOTVVSSTENCAKSTVTTTTVTKLSTPS 1500
Db 13700 QATSVISELISL-----SVQEVSVKESPEDFSDKPKTEKATSIILSEQESISVOEISVKE 13754
QY 1501 TGGSVIILSVK--EOSTVTVVTVTVTDSLTTTGTL-----VISM-----T 1538
Db 13755 ABEVSVEVVKPKTEQATSVISPHM--SLTVQEVTVKGVPADISDVKPKSEKATSIILSEQES 13812
QY 1539 VS--KEVSTDDKYVLMKFSRPKTRSGTALPYRKFTVKTSTKKSIFLVPDDLKLARKGC 1597
Db 13813 ISVQEVSVKEAPGSMKDAKPK--TEQATSV-----ISPHESL--TVQEVVS 13853
QY 1598 IREVPFENNNAKPAIDIMWYPSRPFGITWYRIQTVYKSLAGVSLMLRLMLASLRMDM 1657
Db 13854 VREVPFETISDKP-----KTEKATSIILSEQ-----ESISVOEI 13886
QY 1658 AAKVPPG--GSTRTESETELTTTEIKRRDVGYPYGRFPCYLRK----- 1701
Db 13887 SVKEAPGSMKDAKKTQATSVISPHESLTVQEVSVKEVPEISDKKPKSEKATSIILSEQ 13946
QY 1702 ---IICPIGVPTPK-----ETPTPQKGLRS--SALRPRKPTPK 1737

```

```

Db 13947 ESISVOEISVKEAPESMKDAKPKTEQATSVISPHESLTVQEVSVKELPEISDKKPKTEK 14006
QY 1738 QTGVIILETVWAEELLELMEIRFAERVEREKAQAVQAKRRLQOKPTVIATSTSP-- 1796
Db 14007 ATS--ILSE-----QESISVOEISV-----KEAPGSMKDAKPKTEQ-----ATSVISPH 14048
QY 1797 -----TSSTTSPISPAQVMVAPISGVTGTAKMVLTTKVS 1833
Db 14049 ESLTVQEVSVKEVPTDISETKPKTEKATSIILSEQESISIOEVSVEKATG-----SKMDAK 14103
QY 1834 PATVTFQONKNFQTFATWVKQOSNGVVOVQOKVLIIGIPSTSGOQFTSFQRTAF 1893
Db 14104 P-----KTEQATSVISELIS--LTVQEVSVKESPEDFSRKKTET 14142
QY 1894 VTIKPNISGSGGTTNSQVITGQPIRGMVITRPLQOSTLGAII--RTPVWQPGAQ 1951
Db 14143 AT-----SILSEQESISVOEISVKEAPCSMEVVKPKTEQAT--TILSHESLTVQEVAVK 14195
QY 1952 QVMTQIIRGQF--VTAVASPNTVSTPGOKSLTATSNSIQSSASQPPRPGQOVKL 2008
Db 14196 EVPTDISDKPKPQTOQATSVFSP-----QESISVOEVSVEKAPGAVDSKPKTEHAKS 14247
QY 2009 TMAQLTQLQHGNGOGLTV-----VIOGQGGTTGOLILPQGVTVLRPGQOLMQAAM 2062
Db 14248 TISPLESLTV-----QETVKESPRDIIEELKPKTEKATSVILSEQESI--SVEELIVKQV 14299
QY 2063 PNCTVQRFLETPLATATATATSTTTTVSTTAAGTGQROKSKLSPOMQVHODKTLP---P 2118
Db 14300 PGSMDLKPKTEQATSVISPHOSTVIEITVKEETPDLADE--KPKTEIAKSLILPODSA 14358
QY 2119 AOSSSVGAPOKAPQOTAPQ---SARPOQOTQOSPAQPEVOTQPEVOTQOT----- 2165
Db 14359 VQEVSVKEAPVDIKELKPKSEKATKAPVSTQ--ESLSVEVQVGOQVQEMTPQOISETARV 14417
QY 2166 -----VSSHV--PSEAOPTHAOSSKPOVAAQSPQSNVQGSFVRQSPS--- 2208
Db 14418 SVTKRDSVOVLEVSSBELADIDIPETRLIEDKAKVVTENILQRAQD--EDVIYVLRGPTKEK 14476
QY 2209 -----QTRIRPSTPSQF-----SPGQOSQV-Q 2229
Db 14477 IIEVESDEBELVKEIPEVKEEIEVEQFKLKPRTKKKKEITTEVYIIRKEKEVIE 14536
QY 2230 TTTSGPPIQPHTSLOIPSGOPOPOVQOSTQ--TLSSGQTLQVNSVSPSR----- 2281
Db 14537 EVIKRPEDEIEVIEVWIKKPKKTEVTIEVEVEVTIIRKPEVIEVYVKKPEKIEBIT 14596
QY 2282 POLOIQOQPOVYIAPPOLQOQVQVLSQIOSQVVAQIOAQOSVPOQITLOLPIQIOOSSA 2341
Db 14597 EEVYIIRKPEEKV--VEEITTEVQIIRKPKKPKPEKIEE-----VTEEVYIIRKPK--EEK 14645
QY 2342 VQTHOIQNVTVQAASVOFOLQROVQOLRDOOQKQKOQI--EIKRPHTLQASNOSEIIOQ 2400
Db 14646 IAEVYTEELVYAKKAEKVEVYEEVQIIRKPKKPEKPEIEVTEVYIIRK--KPEEVYEE 14702
QY 2401 VVMK-----HNAVIE-----HLKQKSMTPAER--EENORMIYONQ 2434
Db 14703 VYIKRPEEVYHEVSEVYIIRKPEKVEEYVTEVYQIKKPKPKPEKIEVTEVYIIRK 14762
QY 2435 VMYITLIDIXDEEQAQAKKRR-----DESVQKRSKQNAKL-----SALKRKHQOL 2483
Db 14763 PEEKVEEVEVYVYIIRKPKKPKPEKIEVTEVYIIRKPEKVEEVEVYIIRKPEEKV 14822
QY 2484 RAEILKRRALLDKDLOI--EVOEELKRLDKLTK-----EKDMLQOLA 2522
Db 14823 HEVSEVYIIRKPEKIEVYIIRKPKKPKPEKIEVYIIRKPKKPEKIEVYIIRK 14882
QY 2523 QATAVAP-----CPVTVPLPAPPA-----PPSPPPPGVQHTGLSTP 2563
Db 14883 EEVYIRKPEEKVEEVEVSEVYIIRKPEKVEEVEVYIIRKPKKPKKPEKIEVYIIRK 14938
QY 2564 TLVPAQOKRKRKEE-----KQSSSKSKKKKMTSTTSKET--K 2598
Db 14939 EVYIIRKPEKIEEITEVVOIIRKPKKPKPEKIEVYIIRKPKKPEKVEEITEVYIIRK 14998

```


QY	1494	TLTSLPNSGSDVITISVEOKSV----	VTTTVYDLSL-----	TTGGLTVS-----	153	
Db	3071	AKPTTSSKNGDLPPSRKRPRLISPIPKPTTIANSTLTSOAEPPVSSGTVISSALA			153	
QY	1537	-----	-----	-----	154	
Db	3131	TPTTSTAAGVSAABGLDNPSTPSAQQCKKSPETPEFDQDLDRISFSAVOSISAEFNS			154	
QY	1546	-----	-----	-----	157	
Db	3191	TSLLDNIADERKIPVASPRATKPLDLYEESKSKVVTISOETESAVSALIGESFTSST			3250	
QY	1580	-----	-----	-----	161	
Db	3251	DYSLDGMDEMSVNLTEPLTVIAEPDEBALAKALETGEBASTLEER-----	EMEPER	3300		
QY	1613	DIWPPSP-----	RPFTGITWRYRL-QTVKSLAGVSLMLRLMASLRMDMAKVPBG	1665		
Db	3307	EAEPRPDEAEIESEPPVVEVLDPEELNKAVOSLKHEEDM-----	DIKADTPQSE	3355		
QY	1666	GSTRTESETETLTTEILKRDVSPYGIREYCIKRICIGYPERPKETPPYOKGLRS	1722			
Db	3356	RDLQIDT-DTEENDE--ADSSCP-----	SLKIDELVQSSSSPEKSTISNN	3397		
QY	1726	SALRPKREPKOTGPVLIETWAAEELELMEIRAEVREKEKAQVQAQAKRLDEOK	1785			
Db	3398	S-----	PTPRETANIDI-----	PWESQPKLSNSTPQSVITLK	3433	
QY	1786	P-----	TVIATSTTSPSSSTTSLSPAQKVMAPIRGSV-----	TTGTAMVLTKVGSFA	1833	
Db	3433	PLDTPKTVPAGLPPSPVKLEPPTTSLDLPQVQVGLPABHSGSISANSVINIDL	3492			
QY	1836	TVTPOQKNFQHTPATWKGOSNSGVVQVQKVLGIIPSTGTSGQUTSPOR-RIATV	1894			
Db	3493	SNVISCSCNTSAASATASASASISFGSPITASQAM--	QASIPRKGPRTTPQOARTQSL	3549		
QY	1895	TIRPTSSSGTSTN--	SOVITGPQI-----	PRG-MTVIRPRLQOSTL--	GKAIIRF	1941
Db	3550	TMOPRTTISPEOTNFAVPMVLSPOSHNPQDGTGTVAGIIRASPHSPHSGRCVQASR			3609	
QY	1942	-----	PVWOPARQOVMTQIIRGPRVSTAVASPNVTSPPGQKSLTSATSTNI	1991		
Db	3610	LVGQLSPYGRMVSQR-SPOQ--	QVQDQOQALITTSQSSNISPLASPTTVLSSNS	3665		
QY	1992	QSSA--	SQRPPOQGVKLTMAQITQLTQSGHSGNOGLTVULQOQOTQQLDLPQST	2048		
Db	3666	PYTTSKVUSYORRMO-----	VPQORSPKSAVEQVOTRQMLT	3701		
QY	2049	VLPSPGOOLMAMPNCTVQRELFTPLATPATSTTTTTSVTGAAGTGEOROKSLSPOM	2108			
Db	3702	I--	PLQKMPRIQVPRNHPT--	ITSKVUTVQROAQDOSVASSPRLGS-----	LPBHK	3748
QY	2109	QVNHQDKTLPPAOSSSVGRAKAPOT-ADRSAPOROTQOSRADEUTQOREVUTQTVS	2167			
Db	3749	NVH-----	LMAHON-----	QOQPVLAKTMAHNOOQNMQFHHQMIQROQNMQ-----	3792	
QY	2168	SHVSEADPTNHAOSKRYALQASQPSNVQOSRYKVSQSTYKIRPSTYOLSFGQ-QS	2266			
Db	3793	-----	QOOLHGSO-QITSARQNMHONQAOQOQNNHONQNLQOQNLQOQNHQNTQON	3844		
QY	2227	QVQTTTSQPIRTPHNTSLQIPSGQFQSOQVQVOS--	STQTLSSGQTLQNVSVSSPSPRQ	2283		
Db	3845	QAOQOQFMO-QIOOHOSQO--	QHVOQOQNAQOQNLISQOQNHOSQOQLNO--	QHQAQO	3895	
QY	2284	LOIQDQPR-QVIAVQLOQO-----	VOVLSIQSQVVAQITQAOQSGVPRQOIK	2329		
Db	3896	QOQDQOQLOQLOQMGRQOQKSPQGVGNLIGSTSLFASQOQHNSQLPFR-----	GVPRQO--	3947		
QY	2330	LOLPIQIOQSSAVQTNQIQINV-VTVQASVQSOLOLR--	QOQLRDQOKKKQOQOILEIKREN	2386		
Db	3948	QHNPQOLSHSPCKRMTVLVSNGVQVQPRAL--	ILRVSSHQPRQOQOQLPQOQSSGSHPR	4003		

QY	2387	TLQSNQSEIIQOK-----VYMKHNAV-----ELKQKSMSPAREEENQRMIVQMYKIL	2440
Db	4004	QKQSLSPGAINPLQTPPLVIAQNTPKIIVQOHIAQNVQPPPTQGNALHYQNO--GKQST	4062
QY	2441	DKIDKEEKOAAKKRRKEEVEQOKRSKQMATKISAL-----LFKHKBOLRAEILKKRAL	2493
Db	4063	PGHVEPTPAMSAQKTSSESIVRTPTPTGLAVISANTVGSILTEENILKISQPKQDEL	4122
QY	2494	LDKDLQLEVO-----BELKRLDKIKK-----EKJLMQLAQTVAAAPCPVPYVLPAP	2541
Db	4123	IEQSK--EVSQDYSAAEVAINDIVIKKLDTPILASKAKAAVYMQAI-----	4167
QY	2542	PAPRPPPPPPGVQNTGLSTPTPLPAVQOKRKREEQSSSKSKAKM	2589
Db	4168	-AAPRIINPQRGND--SMAQETLPLPTTSSVUNNSNDHDHDEDETEFTRQL	4212

ID	09VPL2	PRELIMINARY:	PRI:	5533 AA.
AC	09VPL2:			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Spn protein.			
GN	Spn OR CG18497.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY.			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Vendell M.D., Zhang Q., Chen L.X.,			
RA	Bratton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abbil J.F., Agbayani A., An H.-J., Andrews-Plamkoci C., Baldwin D.,			
RA	Ballew R.M., Basu A., Bakendie J., Bayraktaroglu T., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Gang N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Haveron D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howard T.J., Wei M.-H., Ibegwam C.,			
RA	Jajjala M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kohira C.D., Kraft C., Kraivit S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Lin X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Mekulov G., Mishina N.V., Mobarry C., Morris J., Mosher A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Strivast R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassarman D.A., Weislock G.M., Weissbach J.,			
RA	Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;			
RT	"The genome sequence of <i>Drosophila melanogaster</i> ."			
HL	Science 287:2185-2195(2000).			


```

Db 3667 LVGLSPVSGPRWSP--SPQO---QVOQOTQOAHALITSPSSNISPLASPTTIVLSSSNS 3722
QY 1992 QSSA---SOPPRQOGGVKLTMAQLPLQLGHHGNGGLTVVIGGGOTGOLDLPGCVT 2048
Db 3723 PTTSKVNSVYPRNQ-----VPOSPKSAVEVOTPTQLMT 3758
QY 2049 VLPFGQQLQOAMPNGTVGRFLPTPLATTATSTTTTSTTAAAGTGROSKLSPQM 2108
Db 3759 I---PLOGKMPRIQVPHNPT---ITSKVTVYVPOQATQSGVASSPRLGS-----LPPHK 3805
QY 2109 QVHODKTLPPRQSSSVGRPAQAQPT--AQPSARQPTQROPSPAPQREYOTQPEVOTQTVS 2167
Db 3806 NVH---LNAHQN-----QOQPVYAKMTAHQHQQHMQOQFMHQOQIQRQOHMQ----- 3849
QY 2168 SHVSEAPRTHAOSKRPVAAOSOPSNVYGSPVRYOSPOTRIRPSTPSOLSPGO-QS 2226
Db 3850 -----QOQLHGSOQ--QITSAPOHMQHQAQOQOQHNNQOHLMOQLHAQOHPQOKH 3901
QY 2227 QVOTTTSQPIPIQPHISLQIPSGORPOPOVOS---STQTLSSGQTLNVSVSSPSRPQ 2283
Db 3902 QAOQOQFNQ--QIQOHQSOQ---QHVOYQOQOQOQHLSSQOQHGSQQLNQ-----QHQAQ 3952
QY 2284 LQIQOQQR--QVIAVPOLOQO-----VQVLSQIQSQVVAQIQAQSGVPQOQK 2329
Db 3953 QQLQOQIQKLOQMHGPOOQOQSPGVGHGSGTSIFASQOHNQOLPAP-----GVPOQ-- 4004
QY 2330 LQLPIQOOSAVQTHQIQNV--VTYQAAVYEQLOQRY--QQLRQOQKQKQOQIEIKREH 2386
Db 4005 -QHPOQLSHSSPCRPMTIVSVNGVQPPAL--LTRYGSHSQPNQOQOLPHOQSSGHHH 4060
QY 2387 TLQASMOSEIQOQ---VVMKHNAVI--EHLKQKSMTPAREENQRMIVCQNVKYYTL 2440
Db 4061 QKQLSFGANLPLQTLPLNYIQMTPKIIVQOHIAVQNVPPQPGQNAIHYPQNO--GKDS 4119
QY 2441 DKIDKEKQOAKKRKEESEVEQSKQONATKLSAL-----LFRKKEQLRAEILKKRKL 2493
Db 4120 PRCHVPTPTAMSAQKTSSESVYKRTPTPTGLAVISANTVGSILTEENLIKISQPKODEL 4179
QY 2494 LDKQLOIEVQ-----EELKRLDIKIK-----EKDLMOAQATAVAAPCRPVPLPAP 2541
Db 4180 IEQDSK-EVDSQVWSAKEVNIDSVIKKIDTPLASKDKAKRAVEQAL----- 4224
QY 2542 PAPPPPPPPGVQHTGLSTPLPAPASQKRRKEEKSSSKSKKKM 2569
Db 4225 -APAPIPNPQPGNO--SMAQETALPTTSMNVNSNDHDEDETEFQOL 4269

```

RESULT 14

ID 0906C3 PRELIMINARY: PRT: 5533 AA.

```

AC 0906C3:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE SPEN RNP motif protein long isoform.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOL;
RA Wielelte E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
RA McGinnis W.;
RT "Spen encodes an RNP motif protein that interacts with Hox pathways to
RT repress the development of head sclerites in the Drosophila trunk.";
RL Development 0:0-0(2000).
DR EMBL; AF188205; AAF13218.1;
DR HSSP; P09651; IHA1.
DR FlyBase; FBgn0016977; spen.

```

```

DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PSS0102; RRM; 3.
SQ SEQUENCE 5533 AA; 597114 MW; AFG60606DDEF269 CRC64;

Query Match 3.9%; Score 560; DB 5; Length 5533;
Best local Similarity 19.2%; Pred. No. 7 le-14;
Matches 520; Conservative 379; Mismatches 931; Indels 878; Gaps 118;

QY 325 DRSRRKYWFLNRLIIEEDTENENEKIIWYSTKYVLAELIDCLDKYWEALCKLIEEM 384
Db 1997 DKQEKKEKIREKDLREKEQRENDREK-----ELRQKDLREKEM-REKEQR 2041
QY 385 REEIIHNMITE-----DLTKAGSKNSFLAANEELLESIRAKKGDIDNV 431
Db 2042 EKELHREKQREHREKQSRAMQVEQGRGM-----RELSTYQSKMDIAGE 2093
QY 432 KSPFEETKDNENETENSQKAEKNREEFEDQSL--KSDDKTPDDDPQEGKSE----- 482
Db 2094 AS-----SLTADICQHNKENAMDTIAQGTGASPTSPSDTPEKERSKLRSNP 2142
QY 483 VQDFSEKNGELSESPPGAKGASGSTR-----ITRLRNPDK-----LSQLSQGV 530
Db 2143 VRLHKRRLSSQESNHSAGCGGSGSSHQIHNEQYKRLIMENSONISVHSGNRLNDR 2202
QY 531 AAANHAANKLFKEGKQEVLYVNSQGEISR-----LSTKK----- 563
Db 2203 DSKEHKSS-FREDK-----NSSHISPHGCGGSSASSSKHHNRDKHHQGSASSTET 2256
QY 564 ---EYIM-----KGINNYFKLGQ-----EGKYRVYNYQSTNSFALN-----HOHR 603
Db 2257 NSIEVYVPIQOTHNNTSEELQSHQPKREKREHNSHANSSSRHKSKRPHNHR 2316
QY 604 -----EDHDKRHLAHKFLTPAGEFKWNGSVHGSVLTITSLRLITTOLE 649
Db 2317 EKRHSVAESTWIDEHHPQOHNPHR-RISAAG-----GSSAG-----ELSSAA 2359
QY 650 NNIPSEFLPHNASHRAWMIKAVQMSKREFALALILECVKRVVPLPIWRELHGTFR 709
Db 2360 TITSSGKLHHQ--HHRVSERKSSGSDGHHSSKSL-----RAK 2398
QY 710 LHRMTSTIEREK-----EKVKKKEKQEEETMOATVVKYTF 747
Db 2399 LMLSSADSDPTDASKKHSIFDIPDCPNVSMYKVARQSKNNQRAEKKIKAKFS- 2457
QY 748 PVKHQVWKQKGEYRYVYTGQMS---WISKTHVYFVFKLPGNTVNVYKSLGEGTKNMD 804
Db 2458 ---QKOSRAKKRSTSYDGDSDTEFEDRQH-----RNSGSSSFHGR--YFGLSSSD 2505
QY 805 ENMDESDRKCSRPKIKIEPDSEKDEVKSGDAKAGADQNE-----MDISKITEKKQOD 859
Db 2506 DDEDETHQRIS-----SSDAFHGQDQAGASTLADANRVRQMQ-QN 2547
QY 860 VKELIDSDSKPCKEPEVDDMKTESHVNCQESSQVYVNVSEG-----PHLRTSY 912
Db 2548 LRLICDGDSDS---SED--EIRRVKMHSHFGKRNNSNTRIASDESQSPAPADLTIQEH 2602
QY 913 -----KKTKSSKLDGLERIKQTLLEKQRLKILEGGIKGIGKSTMSKNLS 964
Db 2603 PIAPQEIIRQLSDQEKFKSRHDSNSIEER---KLKTEREI---KTELQDFYNS 2654
QY 965 ESPVITKAKEGCQSDSMQEOQSPNANNQPPDLQCGSQSDSS-----YLRMSDPG--- 1015
Db 2655 EYITGKIKLEISPETRKKHKKSKRR-----IKSSSTADTSAQAQPLVWTPLTPTSTFD 2706
QY 1016 -HTTNKILPKDRLVDVSI-----RSPETKCPKONSITNDIEEKVSDLAS 1059
Db 2707 VHSSECKTKFDNPDLDLKECCSITPLEISAGERKHKERKKKREKILRNMTAEAVPN--- 2763
QY 1060 RQGEPTKSKTGDNFFIDSKLASADIGITLCKNKKPFLQOESDITVYSSSKSLHSSVP 1119
Db 2764 ---SPTTNDT-----SSKLSKEERHRLKSKSKSKSMDSNCSNTKIYNSGA-H---P 2808

```



```

QY 1120 KSTNDADATPLSR-----AMDEGKLGCDSESNSTLSSDVIISIODSSEEDM 1167
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2809 STSPSLPATPTAPSPAQTGSKRGEDKMEFIIGJISDEESQFPEQAEFTNKDILPSS---- 2864
QY 1168 IVQNSMESISDEQRTEDQVEVLEPLKCELVSESTGNCEDRLPVGTEANGK-----KPS 1223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2865 -VSTGPIVSAALQTYKQ-----EP-----STPNS-----KNEAHIDLTVHEPE 2903
QY 1224 QOKLEERPVNKCSDQIKLKNITDKNNENRESEK-----KGORTSTFOINGKNKRTIYL 1279
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2904 QOQOQLER-----SRISGSSSSSHADRRHRREKREKREKREKSOEONQIHOKSSK----- 2955
QY 1280 KCECLKEISERVSGNVSEPKVNNIKIIPENDIKSLTVKESAIRPFIINGDIWEDFENER 1339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2956 -----VETKVDDNSVDMDEAGRALEAQ-----LMSDFDK 2986
QY 1340 NSEETKSHLSSSDAAGNRDLSLETLPSTR-----ESDSTQTTTSPASCPRESNNOV----- 1392
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2987 PTSE-----EATPSTPAATYRSDMTDVFPRSDN-EDNNSVDMTKQGY 3026
QY 1393 --EDMEIETSEVKKYTSPTITSEESNLSDNDFIDENGLPINKNENNGESKRKRTVITEVT 1450
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3027 KSPQOQOHKSKOKKKKKKKKEKOKELL-----QOQRESLPNVA 3067
QY 1451 TMTSTVATESKTVIKVEKDQOT-----VNSTENCAKSTV-----TTTTTV 1493
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3068 STSSAPPTGKLTIVNQAAASKHADLQDAKHHSPPYCKPSLPCLIGDDDDALHTPK 3127
QY 1494 TKLSTPSTGSDVILISVKQSKTV-----VTTYTDTSLT-----TTGTLVTS----- 1536
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3128 AKPTTSSKGNGLPISRERKPLISPIKPTITANSSTLSTQSAETPVSSGTIYSSSALA 3187
QY 1537 -----MTVSKEYST 1545
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3188 TTPSTSTAGVSAAPGLDMSPTSASAOCKKSEFIPGPDQLDRISESAVQISAFENS 3247
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1546 -----RDKVKLMKFSRPRKTRSGTALPSYRKFEVTKSTRK----- 1579
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3248 TSLDNIADBPRIKIVASPPRATKPLDLKEESKSRVTISQETESAVSALLGESFGTSST 3307
QY 1580 -----SIFVLPNDLKKLARK-----GIREVYFNYNNAKRAL 1612
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3308 DVSLDGMDSESVNELETPLVLAEPDEEALAKALETAGEPASILEEP-----EMEPEP 3363
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1613 DLMVPYSP-----RPTFGITWYRL-QTYKSLAGVSLMLRLMASLRMDMAKAPPG 1665
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3364 EAEPDPPAEIESEPVVEVLDEPELAKAVQSLKHEDM-----DIAADTPQSE 3412
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1666 GSTRTETSETEITTEIIRRDVGPYGIREFYCIRKIICPIGVETPKETPRPQKGLRS 1725
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3413 RDLQIDT-DTEENPDE---ADSGP-----SLKIDETVQSSSPESKISNN 3454
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1726 SALRKRPRPTPOTGVIIETVWAELELEIMETRAFAERVEKKAQAEVQQAARLEQOK 1785
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3455 S-----PTPRETANIDI-----PNVESOPKLSNESTPOPSYITKL 3489
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1786 P-----TVIATSTSTSTSTISIPAKVMAVAPISGSV-----TTGTMKVLTXXGSPA 1835
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3490 PFLDPTKTYPAGLIPSPVIAIEPTTISKLOOPLQVQYULPAHSHSGSISANSVINL 3549
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1836 TVTFQONKNFHOTFATWVKOGOSNGVVOQKVLGIIPSTSTGTSOOTFTSPQ-RTAVY 1894
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3550 SNVSSCSMTSASAATASASASISFSGPTASQNAM---PQASIPKQGPITPQQAIRTQSL 3606
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1895 TIRPNTSGSGGTTSN---SOVITGPQI---RG--MTYIRTPLOQSTL---GKAIIIR- 1941
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3607 IKQOPPTISIEQOTPRFAVQOMWLSPOSHPQOQGTWVQIRAPSPSPHSPRGQAOSR 3666
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1942 -----PVMVOGAPGOOVVOTIIRGQPVSTAVSAPNTVSTPGQKSLTSATSTSN 1991
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3667 LVGQLSPVGRPNVSP-SPQO---QVOQTOOQHALLITSPQSSNISPLASPTTIVLSSSNS 3722

```

```

QY 1992 QSSA---SOPRRQOGVOKLTMAQLQTLQGHGNGGLVIVIGOGGOTGQOLILIRQV 2048
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3723 PTTSKVNSYQPRNQ-----VPOQRPSPKSAVEVOTTTQOLMT 3758
QY 2049 VLRGPOQLMAMRNCSTVORFLFTPLATTATASTTTTIVSTTAAGTGQRORSKLSRPM 2108
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3759 I---PLOKMPRIQVPHNPT---TISKVVTVQPOAQOSVASSPRLGS-----LPRNK 3805
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2109 QVHNDTLPRQOSSSVGRAPAKOPOT-AQPSARPORQTOPQSPAPREQOTQREVOTQTVS 2167
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3806 NVH-----LNNHN-----QOQPVIAKTAHNOHNOHMOQFMHQMTQROHNO----- 3849
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2168 SHVPSAQPTNHAOSKPRQVAOSQPSNVQGSQSPRVQSPQSTGRITRSTPQSLSPQ-QS 2226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3850 -----QOQLHGQSO-QITSAPROHNOHNOHNOHNOHNOHNOHNOHNOHNOHNOHNO 3901
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2227 QVOQTTSQPIPIQPHNSLQIPSGQOPQSPQVQVQVQVQVQVQVQVQVQVQVQVQVQV 2283
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3902 QAOQOFNQ-QIOHOSQO---OHVOQOQOQOQOHLQOQOQOQOQOQOQOQOQOQOQOQO 3952
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2284 LQIQOPQ-GLIAPVQLQOQ-----VOVLSQIQSQVVAQIQAOQSGVPQOIK 2329
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3953 QQLQOIQKLODMHGRQOQOKSPQGVGHGGSTSIQASQOHNQOLPAR-----GVPOQ-- 4004
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2330 LQLPQIQOSSAVQTHQIQNV-VTVQAAVQEOQLQV--QQLRQOQKQKOQILEKREH 2386
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4005 -QHPOQLSHSPCKRPTLVSVNOGVQPRAI---LTRGSHSOPNQOQQLPRHOQSSSGHPR 4060
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2387 TLQASQNSE-----TIQOVYMKH-----NAVIEHLKOKKSMTP 2420
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4061 QKOLSSPGANLPLPTPLNVIONTPKITIVQOHIYAQONVPPQOTQGNALYHPQONGKQDTP 4120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2421 AEREENQRMIVCNOVMYILDKIDKEKOAKRRKESEYEOKRSKONATKLSAL----- 2475
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4121 RGH-----VEPTPAMSAQKISESVYIARTPTTGLAVISANTV 4159
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2476 --LPRKQOLRAELTKKRALLDKDLQLEVO-----ELKRDILKIK-----EKDLMQL 2521
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4160 GSLTEENLILISQPKODELIEDSK-EVDSDYSAKEVINDISVILKLDTPPLASKDAKRA 4218
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2522 AQATAVAAAPRPVPLPAPRAPRPPPPPPGVQHGILSTPLPLPVASQKRRKEEEDSS 2581
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4219 VEMQAI-----APAPIPNPQGNQ--SMAOETALPTTSMVNSNDHDE 4261
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2582 SKSKKKKK 2589
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4262 DETETROL 4269
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

Q9NHN1
ID Q9NHN1 PRELIMINARY; PRT: 5554 AA.
AC Q9NHN1:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Split ends long isoform.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eurygola; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;
RT Split ends encodes large nuclear proteins that regulate neuronal cell
rate and axon extension in the Drosophila embryo."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF221715; AAF34661.1; -.
DR HSSP; P09651; 1HA1.
DR Flybase; FBgn0016977; spen.
DR InterPro; IPR000504; RNA_rec_mol.

```


DR Pfam: PF00076; tm: 3.
DR SMART: SM00360; RM: 3.
DR PROSITE: PS0102; RM: 3.
SQ SEQUENCE 5554 AA; 599188 MW; 4037E27833D0C622 CRC64;

Query Match 3.9%; Score 560; DB 5; Length 5554;

Best Local Similarity 19.3%; Pred. No. 7, 1e-14;
Matches 520; Conservative 382; Mismatches 948; Indels 838; Gaps 118;

325 DSRRRKWLNRLLIIEEDTENENKTIWYSTKVOALAEILDLDKDYWEALCKLDEM 384
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1991 DKQKEKEIREKDLREKQEREDNREK-----ELRKDLREKEM-REKQQR 2035
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 385 REEIRHNDITE-----DLINKARGSKSFLAANEELIESIRAKKGDIDNV 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2036 EELHREKQDQREHREKQSRRAADVEQGRGM-----RELSSQKSKMDIAGE 2087
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 432 KSPETEKKDKNETENDSDAEKNREFPDQSLK-KSDDKTPDDDPDQSGE----- 482
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2088 AS-----SLTALDCQHNKNAMDTIAQCTPGASPSTPSDNTPKERSRLKRSNP 2136
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 483 VQDFSEKSNGLSESPAGKAGSGSTR-----ITRLRNPDSK-----LSQLKSOQV 530
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2137 VRLHRRRLSQESNHSAGGSGGSHQIHEDYVKRIEMENSONISVHSSNQRLNDR 2136
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 531 AAAAHEAKLFEKGEVLVVSQGEISR-----LSTKR-----563
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2197 DSKERKSS-FKEDK-----NSSSHISPHGCGSSSASSKHHNRDKHNOKGSASTET 2250
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 564 ---EVIM-----KGINNYFKLGQ-----EGKYRVYHNOYSTNSFLNK-----HQHR 603
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2251 NSSIEVVVDPISQTKHNLNTSEELQSHQPKREKEREHFSSHAUSSSRHKSQKDHNNR 2310
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 604 -----EDHDKRRHLAKFLCPAPGEPKMGVSHGSKVLTITLRLITQLE 649
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2311 EKKHSAVESTNTDEHTPOOHNPDR-RISAAG-----SGSAG-----ELSSA 2353
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 650 NNIPSEFLHPMAWASHRANWIKAYQMSKPREFALALALECAVAPVYMLPIWRFLGHTR 709
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2354 TWTSSGKLHHQ-HHRSVERKSSRGSDGHHSSKSL-----RAK 2392
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 710 LHRMTSIEREEK-----EKVKKREKQEEETMQOATWVKTTF 747
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2393 LMLLSASDSDTDASKKHSIFDIPDDCPVSWMDKYKASKCKMMQQAEEKTKAKAFS- 2451
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 748 PVKHQVWKQKGEERYVGYCGMS---WISTHYTRVPKLPQNTVNYRKSLEGTKNMD 804
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2452 ---QLKQSRKKKRTSYDGDSDTEFERDQH-----RNSGSSSFHGR---YFGLSSDD 2499
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 805 ENMDESQRKCSRPKKIKIEPDSEKDEVKSPDAKAGADQNE---MDISKITEKKDQD 859
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2500 DDDDETHQRRTS-----SDSDAHHGQDQNGASTLADANRYRQMO-QN 2541
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 860 VKELLSDSDSKPCKEEMVEDDDMKTESHVNCOESSQVDVNVSEG-----FHLRTSY 912
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2542 LRRLCDGDDS---SED---EIRRVNMKSHFGKRNNSNSTRIASDSESQAPADLTIKQH 2596
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 913 -----KKTKSSKLDGLERRIKQPLTEEKORLEKILLEGIGIGICTTNSSKNLS 964
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2597 PIAPAEIKKEQLSDEQAKKSHDSNSTEER---KLTEREI---KTELGDIFYNS 2648
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 965 ESPVITKAKGCSQSDSKROSPNANNQPEDLIQGCSSQSDS-----VLKMSDS--- 1015
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2649 EYTYTGKLRKYSPETRKKHKKSKRR-----LKSSSTADTSAATPLVMTPLTFSIID 2700
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1016 -HTTNKLYPKDRVLDVST-----RSPETKCPKQNSIENDIEEKVSDLAS 1059
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2701 VHSSECKTKFDNFDLKTECSSIPLEISAGERKKHKEKREKREKLRMWTEFTVFN--- 2757
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1060 RQGEPRKSKTKGNDFTIDSKLASADIGTLCKKPKLLOESDPTIVSSSKALHSVAP 1119
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2758 ---SPTTNDT-----SSEKLSKEERHLRKKSKSKSMDSNCMTKIYNSSGA-H---P 2802
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 1120 KSTNDRAVTPLSR-----AMDEBKLGQDSFNSNTLSSDPTVSIQDSSEEDM 1167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2803 STSPLEPATPTPSABPTAOTSKRGEDKMEFFIGILISDEEESQPPQOAEITNKDIIIPSS----- 2858
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1168 IVQNSNESISQOFRTRQEDVEVLEPLKCELVSGESTGNCEDRLPVKGTENGAK-----KPS 1223
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2859 -VSTGPIVSAALQTYKQ-----EP-----STPNS-----KNEAHIQULTVHEPE 2897
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1224 QOKKLEERPVKCSQOLKATYTDKKNNESEK-----KQRTSTQIQINGKDKPKIYL 1279
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2898 QQQQLER---SRLSGSSSSSHADREHREHREKREKREKRSQROQNOIHQKSSK----- 2949
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1280 KGECLKEISESRVYVSGNVEPVNMINIKIIPENDIKSLTVKKSALRPFIINGDVIMEDNER 1339
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2950 -----VEIKVDDQNSVDMEARALEAQ-----LMSDFDK 2980
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1340 NSSETKSHLLSSDAEGNRYDLSLETLPSTK---ESDSTQTTTPSACSPESVSNQV----- 1392
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 2981 PISE-----EATPSTATYRSDMTDVFREFSDN-EDNNSVDMTQGV 3020
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1393 --EDMEIETSEVKKVTSSPITSEESNLNDFDENGLPIKNENNVGSKRKVTITEVT 1450
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3021 KSEQOEOHKSQDKKKKKRKSKEKQEKLL-----QQQRRRESLPNVA 3061
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1451 TMTSTVATESKTVIKVEKGDQOT-----VVSSTENCAKSTV-----TTTITTV 1493
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3062 STSAPPTPKGLTVNVQAASKHADLDQDAKHISPPCKRSPSLPCLIGDDDDDLHTPK 3121
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1494 TKLSTPSTGSGVDLIYSKEQSKTV---VTTYVTDLSLT-----TTGGLVTS--- 1536
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3122 AKPTTPSSRGNDGLTPSRKPRILSPLPKPTPIANSSTLSTQSAETPVSSGTVIASSALA 3181
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1537 -----MTVSKREYR 1545
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3182 TTPTSSTAAGVSAAPGLDNPSTASAOCKKKEFIEPFGQLDDRISESAVOSISAEFNS 3241
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1546 -----RQVKLMKTSRPKRTSGTALPSYKRVKTKSTK-----1579
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3242 TSLDNIADPEKIPVAPSPRATKPLDKLEBSKGRVITISQETESAVALLGESFGTSSTT 3301
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1580 -----SIFVLPNDLKLKLARK-----GGIREVPYRNYNAKPL 1612
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3302 DYSLDGMDEMSVNELETPLVIAEPDEEALAKAIEAGEPASLIEEP---EMEPEPR 3357
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1613 DIMPYRSP-----RPTFGITTRVRL-QTVKSLAGVSLMLRLMASLRMDMAKVPPEG 1665
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3358 EAPRDPDPEAEIESEPVVEVLDEPELNAKVAQSLKHEDM-----DIKADIPQSE 3406
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1666 GSTRTETSELTITTELIRRDVGPYGINPEYCIRKIICPITGVETPKETPPQKGLRS 1725
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3407 RDLQIDT-DTEENPDE---ADSSGP-----SLKIDETVQSSSPEKISNN 3448
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1726 SALRPKRPETPKQTPGVIIETWVAEELMELIRAFAEVEKEKKAQAVQOAKKRLQOK 1785
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3449 S-----PFRPRTANIDI-----PVNESQPKLSNSTPQPSVTIKL 3483
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1786 P-----TVIANSTSPSTSTSTISPAQVAVARISGSV---TTGTAKVLTITVGSFA 1835
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3484 PFLDTPKTYPAGLRPRVRLIEPTTISKLOQLVQVQGVGLRPHNSGGSISANSVINIDL 3543
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1836 TYTPQONKRFHQTATWVAGQSGNSGVVVOQKVLGIIPSSSTGSOQTFSQPR-RTATV 1894
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3544 SNVSSCSMTSASASATASASASISFSQSPASQONAM---FOASTPQGRITPQQAIRGTSL 3600
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1895 TIRPNTSGSGGTTSN---SQVITGPQI---RPG---MYIRTRPQDQSL---GKAIITP 1941
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3601 IMQPPITSLPEQTPPHANPQWLSFQSHHPQDQGTVMVGIRAPSHSPLHSFGKGVASGR 3660
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1942 -----PYMVPGAPQVMTQIIRQOAPVASTVSNPYVSSPQGKSLTSATSTNI 1991
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3661 LVGQLSPVGRPVNSQPR-SPOQ---QVQQTQOQHALLTSPQSSNISPLASPTTRVLSSNS 3716
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 1992 QSSA---SOPPRPQOGVAKLMAQLTQLTQGHGNGQGLTVVYIQQGQYTTGQDLIPQGV 2048
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |


```

Db 3717 PTTSKVNSTQPRNOQ-----VPOQSPKSVAEVQTPQLMT 3752
QY 2049 VLPBGQOLNQAAMPNGSTVORFLTPLATTATATSTTTTSTTAAGTGEOROSKLSBOM 2108
Db 3753 I---PLQKMTPIQVPHNPT---ITSKVTVQPRQATOSQVASSPRLS-----LPPHK 3799
QY 2109 QVHODKTLPPAOSSSVGRAPAKAPOT-AQPSARPOPTOPQSPAPREVOQOPREVOQOTTVS 2167
Db 3800 NVH---LNAHON-----QOOROVIAKMTAHNOQHMOCFMHOQMIOROOHMO----- 3843
QY 2168 SHVSEAOPTHAOSKPOVAASQSPQSNVQSQSPVRYQSPSQTRIRPSTPSQLSPGQ-QS 2226
Db 3844 -----QOOLHGOSQ-QITSAPQOHMOQHOAQOQOQHNQOHLNOQLHAQOHPTQKH 3895
QY 2227 QVOTTSQPIPIOPNTSLQIPSGQPOSOPQVOS---STOTLSSGOTLNOVSVSSPSRPQ 2283
Db 3896 QAOQOQFNO-QIOQHOSQO---QHQVOQOQNAQOQOHLSQOQHSQOQLNO-----QHQAQO 3946
QY 2284 LQIOQPPQ-QVIAVPLQLOQ-----VOVLSQIOSQVVAQIOAQOSGVPQOIK 2329
Db 3947 QOLQOIQKLOQMHGPOQOKSPQGVGHGSGTSIFASQOHNSQLPAR-----GVPOQ-- 3998
QY 2330 LQLPQIQQSSAVQTHQIQNV-VTVQAASVQEOQLQRY--QOLRPQOQKKKQOQIEIKREH 2386
Db 3999 -QHPOQLSHSSPCKPNTLVSNOGVOPPAI---LTVGSHSQPNQOQOLPRHOSSSGPH 4054
QY 2387 TLQASNOSEIIOKO---VVMKHNVAI--EHLQOKSMTPAERENQRMIVCNOVMKYL 2440
Db 4055 QKOLSSPGANLPLQTPNLVIONTPKIIVOQHIVANOVPPPOTOGNAIHYPONO-GKDST 4113
QY 2441 DKIDKEKQAKKKRRESEVEOKRSKONATKLSAL-----LFXHKQOLRAEILKKRAL 2493
Db 4114 PPGHVEPTPAMSAQKTSSESYSVIRPTPTGLAVISANTVGSLLTEENLIKISQPKODEL 4173
QY 2494 LDKDLQIEVQ-----BELKDLKIKR-----EKDLMQLAQATAVAAPCPPTVPLPAR 2541
Db 4174 IEQDSK-EVDSDYWSAKAEVINDSVIKKLDTPPLASKDAKRAVEMQAI----- 4218
QY 2542 PAPPSPPPPPGVQHTGLSTPLPVASQKRRKEEKSSSKSKKKM 2589
Db 4219 -APAPTPNPQPGNO--SMAQETALPTTSMVSNNSNDHDEDETETROL 4263

```

Search completed: November 20, 2002, 16:33:47
 Job time : 315.851 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:31 ; Search time 22.4905 Seconds
(without alignments)
5128.638 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 1433
Sequence: 1 MYSEEEEDGDAEETDSE.....KLGFKASRSHNNKLQSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3363	23.5	810	1	FALZ_HUMAN
2	441	3.1	3726	1	TRX_DROME
3	404.5	2.8	3256	1	K167_HUMAN
4	401.5	2.8	1367	1	AMNH_YEAST
5	391	2.7	5179	1	MUC2_HUMAN
6	382	2.7	2453	1	NCRL_MOUSE
7	380.5	2.7	3911	1	AKA9_HUMAN
8	379.5	2.6	1589	1	PHP_DROME
9	375.5	2.6	3924	1	ANK2_HUMAN
10	374.5	2.6	2116	1	MYR2_DICDI
11	374	2.6	4377	1	ANK3_HUMAN
12	360	2.5	1875	1	MLP1_YEAST
13	360	2.5	2492	1	ATRX_HUMAN
14	358.5	2.5	3828	1	TRX_DROVI
15	356.5	2.5	2440	1	NCRL_HUMAN
16	355	2.5	2035	1	HFC1_HUMAN
17	354.5	2.5	2349	1	TPR_HUMAN
18	348.5	2.4	1140	1	YMG6_YEAST
19	347.5	2.4	2459	1	MABP_RAT
20	343	2.4	2464	1	MABP_MOUSE
21	337.5	2.4	1790	1	USO1_YEAST
22	337.5	2.4	2090	1	HFC1_MESAU
23	337	2.4	2843	1	APC_HUMAN
24	333	2.3	2468	1	MABP_HUMAN
25	333	2.3	3310	1	CENF_HUMAN
26	331.5	2.3	2230	1	GOC4_HUMAN
27	330.5	2.3	1531	1	NPY5_HUMAN
28	329.5	2.3	3866	1	HRX_MOUSE
29	328	2.3	2804	1	MOZ_HUMAN
30	326.5	2.3	2845	1	APC_MOUSE
31	326	2.3	3562	1	PGCV_CHICK
32	326	2.3	4385	1	VP73_CAEBL
33	324	2.3	2476	1	ATRX_MOUSE

34	322.5	2.3	2414	1	P300_HUMAN	009472 homo sapien
35	321.5	2.2	1411	1	TCOF_HUMAN	Q13428 homo sapien
36	321	2.2	1849	1	IGAF_HAEIN	P45386 haemophilus
37	320.5	2.2	1781	1	AKAC_HUMAN	002952 homo sapien
38	320.5	2.2	3321	1	PCN2_HUMAN	095613 homo sapien
39	319	2.2	3726	1	ABP1_MOUSE	061329 mus musculu
40	316.5	2.2	3329	1	BRG2_MOUSE	P97829 mus musculu
41	316	2.2	1260	1	ALSL_CANAL	P46590 candida alb
42	316	2.2	3969	1	HRX_HUMAN	Q03164 homo sapien
43	314.5	2.2	1616	1	P200_MYCG	Q49429 mycoplasma
44	313	2.2	2869	1	RBP1_PLYAB	000798 plasmodium
45	309	2.2	2441	1	CBP_MOUSE	P45481 mus musculu

ALIGNMENTS

RESULT 1	ID	Sequence	Standard	PRT	810 AA.
AC	012830;				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Fetal alzheimer antigen (Fetal Alz-50-reactive clone 1).				
GN	FALZ OR FAC1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=95347245; PubMed=7621746;				
RA	Bowser R., Giambrone A., Davies P.;				
RT	*FAC1, a novel gene identified with the monoclonal antibody Alz50,				
RL	is developmentally regulated in human brain.";				
CC	-1- SUBCELLULAR LOCATION: DETECTED IN BOTH THE CYTOPLASM AND NUCLEUS OF CELLS IN THE DEVELOPING CORTEX. IN THE ADULT BRAIN, IT WAS SEEN ALMOST EXCLUSIVELY IN THE NUCLEI OF NEURONS OF THE NEOCORTEX. IN THE BRAINS OF ALZHEIMER DISEASE PATIENTS, THE PROTEIN IS LOCALIZED IN A SUBSET OF AMYLOID-CONTAINING PLAQUES.				
CC	-1- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE FETAL BRAIN. EXPRESSION IS MUCH LOWER IN ADULT BRAIN AND IS HIGHER IN NEURODEGENERATIVE DISEASES.				
CC	-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL: U05237; AAA97522.1; -				
DR	TRANSFAC: T04682; -				
DR	GENE: HGNC:3581; FALZ.				
DR	MIM: 601819; -				
DR	InterPro: IPR004022; DDT_dom.				
DR	InterPro: IPR001965; znf_PHD.				
DR	Pfam: PF00628; PHD; 1.				
DR	Pfam: PF02791; DDT; 1.				
DR	SMART: SM00249; PHD; 1.				
DR	PROSITE: PS01359; ZF_PHD_1; 1.				
DR	PROSITE: PS50016; ZF_PHD_2; 1.				
KW	Nuclear protein; zinc-finger.				
FT	ZN_FING 251 298				
FT	DOMAIN 4 9				
FT	DOMAIN 30 35				
FT	DOMAIN 42 46				
FT	DOMAIN 657 660				
FT	POLY-GLU.				
FT	POLY-ASP.				
FT	POLY-ALA.				

SQ SEQUENCE 810 AA: 91799 MW: 77E2C992FE5BE96D CRC64:
 Query Match 23.5%; Score 3363; DB 1; Length 810;
 Best Local Similarity 82.6%; Pred. No. 2.5e-113;
 Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2:
 QY 1 MWSSEEEEDGAEETQSSSEDEDEEMEDDDSDYPREMEDDDDDASYCTESSFRSHST 60
 DB 1 MWSSEEEEDGAEETQSSSEDEDEEMEDDDSDYPREMEDDDDDASYCTESSFRSHST 60
 QY 61 YSSTGRRKPRVHRSPILIEEKDIPLEPKSSDDLAVPNEHIMNVAIEVLRNGETV 120
 DB 61 YSSTGRRKPRVHRSPILIEEKDIPLEPKSSDDLAVPNEHIMNVAIEVLRNGETV 120
 QY 121 LRLSPFRREDCAALVSOEQCTLAEMHVVLLKAVLRREEDTSNTTFGPADLKDSVNSTLY 180
 DB 121 LRLSPFRREDCAALVSOEQCTLAEMHVVLLKAVLRREEDTSNTTFGPADLKDSVNSTLY 180
 QY 181 FIDGWTPEVLRVYCESKREYHVLPRQEAEDYRGVPEKIKVLOPLVDFLTNTARE 240
 DB 181 FIDGWTPEVLRVYCESKREYHVLPRQEAEDYRGVPEKIKVLOPLVDFLTNTARE 240
 QY 241 ELMSEGVLYQDHCNVCRLGLDLCETCSAVYHLECVKPPLEVPEDEMOCVCVAHKV 300
 DB 241 ELMSEGVLYQDHCNVCRLGLDLCETCSAVYHLECVKPPLEVPEDEMOCVCVAHKV 300
 QY 301 PGVDDCAVEIQNNKPYIRHEPTGDRSRKRYWPLNRLLIEEDTENENKIKWYSTKVQ 360
 DB 301 PGVDDCAVEIQNNKPYIRHEPTGDRSRKRYWPLNRLLIEEDTENENKIKWYSTKVQ 360
 QY 361 LAELIDCDKDYMAELCKILIEEMRETHRMIDTDLTNKARGSNKSFLLAAAEELLES 420
 DB 361 LAELIDCDKDYMAELCKILIEEMRETHRMIDTDLTNKARGSNKSFLLAAAEELLES 420
 QY 421 IRAKGDIDNVKSPETEKKDNETENDSKAEKNREEFEDOSLEKSDDKTPDDPEQK 480
 DB 421 IRAKGDIDNVKSPETEKKDNETENDSKAEKNREEFEDOSLEKSDDKTPDDPEQK 480
 QY 481 SE----- 482
 DB 481 SEETEVGDKGNSVANLGDNTNATSEETSPSEGRSPVCLSETPDSNMAEKVASEL 540
 QY 483 ----- 482
 DB 541 PDVPEEPNKTCESSNTSATTTISIQPLNLSNSSSELNSQSSAKRAADDPENGERSHT 600
 QY 483 -----VDPEKSEKNGELSESPGAGKAGSSTRIITRLRMPDSKLSQLKSOVAAAA 534
 DB 601 PVSIOEBIVDFTSEKSTGEISESPGAGKAGSSTRIITRLRMPDSKLSQLKSOVAAAA 660
 QY 535 HEANKLFEKEKEVLYVNSGELISRLSTKKRYIMKGINNFYKLGQEGKRYVHYNOSTNS 594
 DB 661 HEANKLFEKEKEVLYVNSGELISRLSTKKRYIMKGINNFYKLGQEGKRYVHYNOSTNS 720
 QY 595 FALNKHQHRDHHKRRHLAKHFCULTPAGEFKANGSVHGSVLTLSITRLITTOLENNIS 654
 DB 721 FALNKHQHRDHHKRRHLAKHFCULTPAGEFKANGSVHGSVLTLSITRLITTOLENNIS 776
 QY 655 SFLHPNM 661
 DB 777 TSLHPSE 783
 RESULT 2
 TRX_DROME STANDARD; PRT; 3726 AA.
 AC P20659; Q27255; Q27327;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax protein.
 GN TRX.
 OS Drosophila melanogaster (Fruit fly).

CC Eukaryota: Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90192757; PubMed=2107543;
 RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
 RT "The trithorax gene, a trans-acting regulator of the bithorax complex
 in Drosophila, encodes a protein with zinc-binding domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
 RN [2]
 RN SEQUENCE FROM N.A.; ALTERNATIVE SPLICING, AND CHARACTERIZATION.
 RX MEDLINE=95009521; PubMed=7924996;
 RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
 RT "The bithorax complex is regulated by trithorax earlier during
 Drosophila embryogenesis than is the Antennapedia complex, correlating
 with a bithorax-like expression pattern of distinct early trithorax
 transcripts.";
 RL Development 120:1907-1917(1994).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Oregon R;
 RX MEDLINE=96100387; PubMed=8555104;
 RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "Conservation of structure and expression of the trithorax gene
 between Drosophila virilis and Drosophila melanogaster.";
 RL Mech. Dev. 53:113-122(1995).
 RN [4]
 RN CHARACTERIZATION.
 RX MEDLINE=95047388; PubMed=7958911;
 RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
 RT "The Drosophila trithorax gene encodes a chromosomal protein and
 directly regulates the region-specific homeotic gene fork head.";
 RL Genes Dev. 8:2478-2490(1994).
 CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
 WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
 CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
 CC ZINC.
 CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, M31617; AAA29025.1; -;
 DR EMBL, Z50152; CAA90514.1; -;
 DR EMBL, Z50152; CAA90513.1; -;
 DR EMBL, Z31725; CAA83516.1; -;
 DR EMBL, Z31725; CAA83515.1; -;
 DR PIR, A35085; A35085.
 DR HSSP, P20393; 1A6Y.
 DR TRANSFAC: T00850; -;
 DR FlyBase: FBgn0003862; trx.
 DR InterPro: IPR003889; FYRIC.
 DR InterPro: IPR003888; FYRIC_C.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001965; ZnF_PHD.
 DR InterPro: IPR001841; ZnF_ring.
 DR Pfam: PF00628; PHD; 3.
 DR Pfam: PF00856; SET; 1.
 DR SMART: SM00542; FYRC; 1.


```

Db 1916 FLCSLRWSSKBEWKIYEYVTRTIIQNSSSTLALDVGNNYVDHHTNPNRSKVOLGMAOI 1975
QY 1283 CLKELISESR---VSGNVE-----EKVNNINKIIPENDIKSLTVEASAPIRINGDVIME 1334
Db 1976 ARKHSTSLASEFLLENGGTGSEFPNPNSC---VPPDONTNEEPQOQADLLPELKDALFE 2033
QY 1335 DENERSSETKSHLLSSDAEGN---YRDSLETLPTSKES--DSQOTTPSASCPSNSVN 1390
Db 2034 DL-----PHELLDGISMDIFLYDKDIDLFALISEQSKDGIOTAMTS-----N 2074
QY 1391 QVEDMEIETSEVKKYTSSEITSEESNLSDNFIENGELPI--NKNENVGESKRKTVITEV 1449
Db 2075 QANONNOAGAGANSVS---ICDEDTNRNSNTSLG--NGWPNANPYEDAMLSAARN---SSQ 2126
QY 1450 TMTSTVA-----TESKTVIKVEKGDKQTVVSTEMCASTYTTTTYTKLSTPTG 1502
Db 2127 VQMLTTLMPKLDGNSAATAIKRKLK-----NLACGVLETLSSQGRNKKEMATV 2178
QY 1503 GSVDIISVKEOSKIVVTTTVDLSLTGTGTLVTSMT--VSKEYSTRD-----KVKLK 1553
Db 2179 AGV-----SRROSISETSEGVATTSQSVRSKSPFTWSAAKRYEKSREAKMRIMQ 2232
QY 1554 FSRPKRTSGTALPBYRKFVTKSKTSIFVLPNDLKLAKKGIR---EVPYFNNAK 1609
Db 2233 MDG-----VDSITFEFRIISGDENLSTAOFSQGVCKDCRCQCTYRNDAF 2276
QY 1610 PALDIMPSPRPFTGTLWRYRLQTVKSLAGVSLMLRLMLASLRWMDMAKVPVGGGSTR 1669
Db 2277 QR-----HLPSGSPF-----MSNTEESDVS-----GQGMIN 2303
QY 1670 TETSETTEITTEIIRKRDVGYGIRFEYCIRKIICPIGVPEPTEPTEPPOR----- 1720
Db 2304 NATOISASLNELOKOLLANAGLNLY-----LOSATSFPQORLGLSLOFGL 2350
QY 1721 KGLRSSALRPK-----RPEPEKQTPVLIETWAEHELML--EIRAPAEV-- 1765
Db 2351 QGLQQLQLOQPSLGNFELSQPNATANT-----DDLQIYANSLQGLAANLGG 2399
QY 1766 -----EKEKAQAVEQAKRLBOOKPVIATSTTSPUSSTSTSPAKVM 1811
Db 2400 GFTIAPQTVAPAPOLIAVSTNPDGQOFLOIPQTMQATTTPTATYQIOLANTTKIM 2459
QY 1812 VAPISGSVTGTKMWLF--TKVGSFATVFOOKNFHQTATWVK-----QGSNSGVAO 1864
Db 2460 L-----PLTAAKPKLKVATKAQAQAAVAKORLKSQHVPKPAKLPHPQHQOQOQO 2514
QY 1865 VQO--KVLG-----IIPSTGT-----SQOFTSP----- 1887
Db 2515 VOQPTVWGONLLOPOLFFQSSSTQTOAPQIILPQAPQNIISVITGDSGQOPLOYITISIP 2574
QY 1888 -----QPR-TATVTIRPTSSSG-----TTSNS--OVITGPQIIRPM 1922
Db 2575 TAGEYKPOPOPATPFTFLTAPAGATYLOTDASGNLVITTPSPNSGLMLTMOISLOAP 2634
QY 1923 TVIRTPLOOSTL-----GKAIIRPVMYOPGAPQOVM-----TO 1956
Db 2635 OVIGTLIQPTILOGGADCN-----QPSNOQPOLILGTGGSGGLEPATTSPOV 2685
QY 1957 IINGQPV-----STAVSAPNTVST-PG--OKSLTSATSTNIGSASOPPRPQO 2003
Db 2686 IIAQPMYGLETIVONTWSSQOFVSTAMPGLSQNASFSATTTQVQASHIEP----- 2740
QY 2004 GOVKLTMAOULTLOLGHGNOGLTV-----VIOGQGO--TTGOLQILPGCVTVLPBGQOL 2057
Db 2741 IYDLVAGVYVLANITGDASSAGFELMAASVLQOQOTDDTTTOI----- 2782
QY 2058 MOAAMNGTVORELFTLPTATTAATTYVSTAAGGEQROKSLSPQMVHODKTLIP 2117
Db 2783 ----LQNAFQ--FGSVPTSSGASTSDYTSYMWTA-----KIPVYDIKKTNAQA 2828
QY 2118 PMS--SSVPAPAAQPO-----TAQPSARPPQOTOPSPAPQEVOTOPQEVOTQTTVSSH 2169

```

```

Db 2829 KAAGISGVKVPPOPOVYVNVKVLPTISIVTQOSQOVKNSLNKOSGVKKAASGTGTTCGA- 2887
QY 2170 VPSEA-----OPTHAKSSKPOV-----AAGSOPSONNOGOSVPR 2203
Db 2888 PPSIASRPLCKKTMIRPIHKLKPKVMKPTPRVQVONHNSLLQOQOQOQOPOLOQOQIPAV 2947
QY 2204 V--QSP-----SQIRIRPST-----PSQLSPGOQSOVQTTTSQPI---PIQPH 2241
Db 2948 VVNOVPPVTTISQORIPATQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3007
QY 2242 TSLQIPSGOGPOQSOPOVQSSSTQTLSSQOTLNQVSVSPSR----- 2281
Db 3008 VQSFVNEPQALQEQEELANRYOHEFTSSSSSSSSNCSPINWVNPQOQAPSTTSSSTRP 3067
QY 2282 ----PQLQIQOQOP-----QVIAVPOLOQOVO--VLSQIQOSQVNAQIOAOSGVPOQIK 2329
Db 3068 TNRVLPMQOQOEPAFLPISNECPVVSPTPRKPVBOPIIHQHTASVSCKVCAOKSTLSPV- 3126
QY 2330 LQPLQIQOQSSAVQTHQIONVTVQQAASVQELQRYQQLRDQOQKQOQOQIETKREHTLQ 2389
Db 3127 --YEAELKSSVLES--IYDVTMDA--ILEQPVYQSIYTEGLYEKNSPGSKTEQLLL 3180
QY 2390 ASNQSEITQOVVMKHNVALEHLKOKSM--TPAREEN-----QRMIVCN- 2433
Db 3181 QOQOQEQNLQNLVNNQYGLDKHTFQVEPMDTVYREEDLEEBEDEDDESLMKRTSACND 3240
QY 2434 -----QVKYIILDKIDKEP-----KQA 2450
Db 3241 HEMSDSEEPVAKDKISKIINDLNDCAOSIATATMEVDASAGYQOQVEDVLAATTAQS 3300
QY 2451 AKRKRRESEVQKRSKONATKLSALLFKHEQRLAEILKKRALLD-XDLQIYQVEELKD 2509
Db 3301 APTEEFEGALFAVAEAAATYINEMDAH-----VLIDKOLQNGVELELRR 3347
QY 2510 LKIKKEKDLQMLQAATVAAAPCPVTVLPAPAPPPSPPPPPG 2553
Db 3348 ---KEQRTVISOEQOSKAIVP--TAAAPBPPOPIQEKKMTG 3386

RESULT 3
ID K167_HUMAN STANDARD: PRT: 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen Ki-67.
GN K167.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schueter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdts J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdts J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO
CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY
CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL
CC CHROMOSOMES.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A

```



```

QY 1545 TPODKYKIMKESPPKKTSGTA-----LPSTK--EYTKSTKSIPLVLPND 1588
| | | | |
DB 1684 LDSAASLGSKROLTPKPKSEVPEDLAGFLFQTPSHTESMNEKTTVSYASQPD 1743
| | | | |
QY 1589 L-----KLAARCGIREVYFYNNAKPAIDIMPYSPRPFEGITMYRIQTVKS 1637
| | | | |
DB 1744 LVDTPSSKPOPKRSLKADTEE-----EPLAFKQTPSAG-----KAMHTPKP 1787
| | | | |
QY 1638 LAGVSLMR-LTMASTLRMDMAKVPKGGSTRTETSETETTTETILKRDVPGYGR-- 1694
| | | | |
DB 1788 AVGEKDIINTFLGTPVQKLDQGNLP--GSNRLOTREKKALEELT-----GEHEL 1838
| | | | |
QY 1695 -----FECIRKTIIC--PIGV-EPKKEPTTPORGLRSS-----ALRPRKP 1733
| | | | |
DB 1839 FQTPCTDNPDADEKTKKILCKSPQSDPADPHTNPKQPKRSLKADVEEELAFKRLTP 1898
| | | | |
QY 1734 -----ETPKQGPVLIETWAAEELMELELRAE-----AEVE-----KE 1768
| | | | |
DB 1899 SAGRAMHTPKAA-----VGEER-----DINTFVGTPVEKLDLIGNLPKSGRRPQTPKE 1946
| | | | |
QY 1769 KAQAVEQOAKRRLDQOKPTVIATSTTSPTSTISPAQVWVAPISGVTGTGKAVLT 1828
| | | | |
DB 1947 KAKLEDLAGKELFQTPGHTEESMTDKITEVCSKSPQDPVATP-----TSSKORLK 2000
| | | | |
QY 1829 TKVSPATVTEQOKNKHQFATWVKOGQNSGVVQVQKVL--GIIPSTGTSGQTFST 1886
| | | | |
DB 2001 ISLCK-----VGKKEEVLPGVKLTLQSGKQTGT-- 2028
| | | | |
QY 1887 FQPTATVTVIRPNTSGSGT-----TSNSOVI-----TPQIRPGMTVIRTP-----L 1929
| | | | |
DB 2029 -----HRETAGGCKSKAKESAKOMLDBANTGEMERP-----RTPEEAQSL 2073
| | | | |
QY 1930 QOSTLGAIIITPVWVGAPQOVMTOIRGQPVSTAVSAPNTVSSSPGOKSLTATSTS 1989
| | | | |
DB 2074 EDLAGFELQTPDHTTEESTTDKTKI-----ACKSPPEESMDTPTST- 2117
| | | | |
QY 1990 NIGSASQPPPOGQVYKLT--MAQLTQLQHGNGNGLTVILOGQOTTGQLQILP-- 2044
| | | | |
DB 2118 -----RRRKTPLGKNDIVELISALKQLQO-----TTHTDKYPGDE 2153
| | | | |
QY 2045 -QGVTVLPGGQOLM-----QAMPNGTVQ-----RFLF--TPLATTTATAS 2083
| | | | |
DB 2154 DKGINVERETAKOKLDPAASVTSKRGQPRPKKAPLEDLAKELFQTPVCTDKPPTH 2213
| | | | |
QY 2084 TTTTIVSTTA-----ACTGE--OROSKLS--PQMVHODKTLPPAGSSVGPAPAKQOTA- 2134
| | | | |
DB 2214 EKTTKIACRSPQDPVQTPPTIFKPOSKRSLRKADVEESIALRKRTPSVGKAMDTPKPA 2273
| | | | |
QY 2135 -----OPSARPOQOTOPQSPAPQPEVQTPQPVQOTIVSS-----HVP 2171
| | | | |
DB 2274 GDEKDKAKFNGTVPQKLDLPGNLPKSKRMPQTPKE-----KAQLEDLAGKELFQTP 2326
| | | | |
QY 2172 SEAOPTHAOSSKPOVAQOSPOSNVQSGPVRVQSPQTRIRPS----- 2215
| | | | |
DB 2327 GTDKPT--TDEKTKRIACKS--PQ-----PDPVDTFASRKORRKRLRKADVEEFLAIR 2377
| | | | |
QY 2216 --TPS-----QUSPGQOQOVQTTSOPIPIQPHSL-QIP-SQGOPOSQP----- 2256
| | | | |
DB 2378 KRTPSAGKAMDTPKPAVSDKKNINFTVEPTVPQKLDLIGNLPKSGKROPPKPEKALEDL 2437
| | | | |
QY 2257 -----QVQSTQTLSSGQTLNOVSVSSPRPOLQIOQPOVOYAVP-----QLOQOVY 2305
| | | | |
DB 2438 VGRKELFQTPGHTEESMTDKITEVCSKPOQESFTSSSKORLKITPLVKADMEEPILA 2497
| | | | |
QY 2306 LSGI--QSOVAVQIOAQSGVPOQIKL--QLPIQIOQSSAVQTHQIONVVT--VQASVOE 2360
| | | | |
DB 2498 VKRLRTSETTQTHTEPGDSKSIKAFKESPKQILDPAAVSTGSRROLRTKEXARALE 2557
| | | | |
QY 2361 QLORVQVL-----RQOQOK-----KKOQOIELKR-- 2384
| | | | |
DB 2558 DLVDEKELFSAGHTEESMTIDKNTKIPCKSPPELDTATSTKRCQPKTRPRKEVKEELS 2617
| | | | |

```

```

QY 2385 --EHTLQASNOSEIIQOVYAKHNAVIEHLKO--KCSMTPAERENORATVQWAKYIL 2440
| | | | |
DB 2618 AVERLQTSOGSTYTHHEPASGDEG-IKVLKORAKKPNVVEEPRR-----EI 2487
| | | | |
QY 2441 DKIDEEKOAKKRRRESEYQOKRSKONATKLSALFLKHEQORA-----EI 2487
| | | | |
DB 2665 -----RPRAPKEKQPLEDLAGFTELSITSHTQESLQATKATIPCESEPLEV 2713
| | | | |
QY 2488 LKRALDLKDLQLEVOELKRLDKIKKEKDMQLAQ----- 2523
| | | | |
DB 2714 VDTYASTKRHLKTPVQK-----VQVKEEPSAVFTQTSGETTDADKEPAGEDKIGALKE 2768
| | | | |
QY 2524 -ATVAPCPPTVPVLPAPPAPPSP-----PPGQHTGLSTPLLPVAPSKRR 2574
| | | | |
DB 2769 SAKOTPPAASVTSVSGRRRPPAPRESAQLIEDLAGFQDPAGHTEESMTDKTKIPCKSS 2828
| | | | |
QY 2575 EEKSSSSKKKKRKMISTTSKETKDPKLYCICKTPYDESKFYIGCDRONMYHRCVGI 2634
| | | | |
DB 2829 PELEDATSSKRRPRRAQKVEYKEE--LLAVGK-----L 2861
| | | | |
QY 2635 LQSEAEILIDYVCPQOCOSTEDAMTVLPLEKDYEGIKVYLSLQAKMAMPLEBPDPN 2694
| | | | |
DB 2862 TQTSGE-----TTHTDKPEVGEKGTKA-----FKQPAKRN 2892
| | | | |
QY 2695 -DAPDYGVYKPEPM-----DLATMEERVQR-RYVEKLETFYAD 2730
| | | | |
DB 2893 VDAEDVIGSKROPRAKRAQPLEDLASFQELISQTPGHTEELANGAD 2940
| | | | |

```

RESULT 4
 AMYH_YEAST
 ID AMYH_YEAST STANDARD: PRT: 1367 AA.
 AC P08640: P08068: 01-AUG-1998 (rel. 08, Created)
 DT 01-FEB-1995 (rel. 31, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Glucanase 51/52 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucanhydrolase).
 GN STAL OR STAZ OR MAL5 OR MAL5 OR YIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Cosey T., Dear S., Devlin K., Fraser A.,
 RA Goulet S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RC MEDLINE=87194600; Pubmed=3106330;
 RA Yamashita I., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RL STAL".
 RL J. Bacteriol. 169:2142-2149(1987).
 RN [3]
 RP SEQUENCE OF 1-31 FROM N.A.
 RC STRAIN=SPX101-1C;
 RA MEDLINE=89031230; Pubmed=3141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STAZ and SGA genes
 RL from Saccharomyces cerevisiae".
 RL FEBS Lett. 239:179-184(1988).
 CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -I- SIMILARITY: TO S.POMBE SPBC215.13.
 CC -I- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL: Z38061; CAA86176.1; -
CC EMBL: M16164; AAA35014.1; -
CC EMBL: M16165; AAA35015.1; -
CC EMBL: X13857; CAA32069.1; -
CC PIR: A26877; A26877.
CC PIR: A26877; A26877.
CC PIR: S48478; S48478.
CC SGD: S0001458; MUC1.
CC Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
CC Signal; Multigene family.
CC SIGNAL: 1 21
CC CHAIN 22 1367
CC DOMAIN 210 1367
CC CARBOHYD 817 817
CC CARBOHYD 874 874
CC SEQUENCE 1367 AA: 136110 MM: 91C00E2DBD61AA9D CRC64;
CC -----
Query Match
Best Local Similarity 21.2%; Score 401.5; DB 1; Length 1367;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;
QY 1135 DEFG---KACGDE-----SNSTLENSPTVSIQDSSEEDMIVQNSNESISEQFRT 1182
DB 191 DFGFYWMNIDCDNNGCGTSTTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST 243
QY 1183 REQDVEVLPELKLCELVSGSGTNC--EDRLPVKGTSEANKKPSQOKKLEERVPKNCSDQI 1240
DB 244 TSSSTSSSTSSSTTAPATPTTSTCKEKPPTTSTCKEKPPTTSTCKEKPPTTSTCKEKP 296
QY 1241 KLTNTDKNNRESEKQRTSTFOINKDKPKIYLKCELCLEISERVVSGNEPK 1300
DB 297 KKTSTTSCKTCTTTPVPPSSSTTE---SSAPVPPSSSTSSSTSSAPVSSSTTE--- 350
QY 1301 VNNINKIIPENDIKSLTVEKSAIRPINDVIMEDFNERNSSSTKSHLLSSSAEGNYRD 1360
DB 351 ---SSAPVP---TPSSSTSSSAPVTS---STSSSAPVTSSTSSSAP 394
QY 1361 SLETLSTKESDS---TQTTTPSACPESSNVQVEDMELETSEVKKVTSSPTSE--- 1413
DB 395 VPPSSSTSSSAPVTSSTSSSAPVTSSTSSSAPVTSSTSSSAPVTSSTSSSAPVTSSTSS 453
QY 1414 -----EESN--LSNDFIDENGLPINKNNVNGESRKRTVITEVTMTST-VA 1457
DB 454 SSAPVPTPSSSTSSSAPVTSSTSSSAPVPTPSSSTSSSAPVTSSTSSSAPV 513
QY 1458 TESKTVIKVEGDKQTVSGSTENCAKSTVTTTITVKLSTPSTGSGVDLIISVKEOSKTV 1517
DB 514 TSSSTSSSAPVPTPSSSTSSSAPVTSSTSSSAPVPTPSS---STSSSTP 568
QY 1518 VTTTVTDLSLTTCGLVTSMTYSKE--YSTRDVKLAKFRPKTKTSAGTALPSYKRFVTK 1575
DB 569 VTSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVTS 628
QY 1576 STFKSIFVLPNDLKLARKGIRVFPYFNNAKPALDIMPYSPPTPGITWRYRLQTV 1635
DB 629 STTESSA-----PVTPSSS-----TTE 647
QY 1636 KSLAGVSLMLRLMASLRMDMAKAVPPGGGSTRTEETSEITTT-----TELKR 1685
DB 648 SSAPVPTP-----SSSTSSSAPVPTPSSST-TESSSAPVTSSTSSSAPVTSSTTE 701
QY 1686 RVGPGVGFIREFCIRKILICIGP-----ETPKETPTQKRKLBSAL----- 1728
DB 702 SSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVTSSTSSS 761

```

```

QY 1729 ---RKRRETPKGTGVIETVWAEELWEIRAFERVEKEKAQAVEQAKRLREQ 1784
DB 762 APVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSSAPVPTPSSSNTSS 821
QY 1785 KP--TYIATSTT-----PTSSSTSTISPAOKVNAVAPISGVTTGKVLTKYGS-- 1833
DB 822 APSTPSSSTSSSAPVPTPSSSTSSS-----APVSSSTSSSAPVPTPSSSN 874
QY 1834 -----PATVTFQGNKNFHOEFATVWVGQSGNSGVQVQKVLG-----IIPS----- 1875
DB 875 ITSSAPSSIFSSSTTESFST-GITVTPSSSKYPGSGTETSVSSTETTVPTTTSVTT 933
QY 1876 -----STGTSG--OTFTSFPQRTAVTIRPTSGSGGTTNSQVYTPQIIRPCM 1922
DB 934 PSTTTTITVTCSTGTNSAGTTSGCSPKVTITV-PTTTTVTSVTSSTTIT-----T 985
QY 1923 TVIRTPLO--OSTLG---KAIIRF-PVMVQPCAPQOVMTQIIRGQPVSTAVS-----AP 1970
DB 986 TVCSTGTNSAGTTSGCSPKVTITVTPCSTPSETASESTTSPPTPVTVVSTTVTTE 1045
QY 1971 NTVSSPFGOKSLNASTSTNIOSS--ASQPPRPQGGV-KLFMAOLTQLQHGNGGLT 2027
DB 1046 YSTSTPGE--ITTTVTKNIPPTTYLTAPTPSVTTVTFPTTTTTCSTGTNSA-- 1102
QY 2028 VVIQGGQTTGLQLIPQGVTVL---PGPQQLMGAAMPNGTVORFLPTLATATTAS 2083
DB 1103 -----GETTSGCS--PKVTTVTPCSTGTGEYTTA-----TTLVTTA----- 1138
QY 2084 TTTTTVSTTAAGTGEOROSKSLPOMOVHDKTLPRAOSSVSGAKKAPQIAPQSPARPOQ 2143
DB 1139 -VTTTVTTSTGTGNSAGTTTG---YTTKSVPTVTVTLAPS----- 1178
QY 2144 TQPSAPQPEVQTOPVQTOPVTVSHVPSFAOPTHAQSSKPOVAQSQPSQVNOGQSPVR 2203
DB 1179 ----APVTPATNAVPTTITTECSAATNAGETTSCSATTIYSSASAGNTAPSAF--- 1231
QY 2204 VQSPQSTRIRPST--PSQLSPGOQSOVQTP---TSQPIPIQPTSLQIPSGQPOQSPQV 2258
DB 1232 --TPVTTAL-PTTVITTESVSGTNSAGETTTGTTTSPITTYITTL-IPSNACKNVEY 1287
QY 2259 QSSQTQLSSGQTLNOVSVSSPRQLOIQPOPOVIAVPOLOQOVVLSIQSQVVAQIQ 2318
DB 1288 ATAT-----NPISIKTTS-----QLATFASASVAPV 1315
QY 2319 AQQSGVPOQIKLOPIQIOQSSAVQTHQIONV--VTVOAAS 2357
DB 1316 TSPS-----LTGPIQASGASAVATYVSPSISTYOGAA 1348

```

RESULT 5

MUC2_HUMAN

ID MUC2_HUMAN STANDARD; PRT; 5179 AA.

AC 002817; Q14878;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Mucin 2 precursor (Intestinal mucin 2).

GN MUC2 OR SMUC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Intestine;

RA MEDLINE=94132002; PubMed=8300571;

RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Sidiki B., Kim Y.S.;

RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.

RT Identification of the amino terminus and overall sequence similarity

RT to prepro-von Willebrand factor.";

RL J. Biol. Chem. 269:2440-2446(1994).

RT [2]

FT	DOMAIN	2277	2281	CORR. BOX OF ID2.
FT	DOMAIN	58	64	POLY-GLN.
FT	DOMAIN	593	602	POLY-ALA.
FT	DOMAIN	1044	1047	POLY-PRO.
FT	DOMAIN	1713	1718	POLY-ALA.
FT	DOMAIN	1968	1979	POLY-SER.
FT	VARSPLIC	2333	2371	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	1952	1952	I -> T (IN REF. 2).
FT	CONFLICT	2090	2050	A -> P (IN REF. 2).
SO	SEQUENCE	2453 AA;	270640 MW;	5220B840382FE6A CRC64;

Query Match	2.7%;	Score 382;	DB 1;	Length 2453;
Best [local] Similarity	18.9%;	Prod No 1	40-06-	

Matches	434;	Conservative	302;	Mismatches	782;	Indels	782;	Gaps	108;
---------	------	--------------	------	------------	------	--------	------	------	------

QY	298	HKPVGTVDCVAELQKKNPYIRH-----EPICGRSKRKWFLNRLLIIEED-----	343
Db	378	HEISEITIDLSJOENNEKOMROLSVTPMMDADAEORVKFTINMGMLDEPMKVKYDROQM	437
QY	344	---TENENE-----KKIMYSTVVOJAEILIDCLDRKYMVAELCTILEEMREEIHR	390
Db	438	NWYTDHEKEIFKDFQIOPHKNNGLASYLEKRSVDCVLYYY-----LTKKNENKALVRR	493
QY	391	IHDITEDLTNRKGRSNGKSFLAANEILIESIRAKGDIIDNVKSPETEEDKNETENEDSKD	450
Db	494	NY-----GKRRGRNOOIARPSOEKVEEKEEDKA-----EKTEKKEEKKDEEK	538
QY	451	AENKREEFDOSLEDSDDKTPDDPECGKSEYDDFKSEKNGSLSSPAGGASGSTR	510
Db	539	DDK--EDSETTKKEKDRTEATA--EPPED--KIOQVTPRGKKTAN-----SOGRGGR----	584
QY	511	ITIRLEBNPDSKLSQLKSQOAAAAAHEANKLPEKEGEVLVYNSOGELISRLSTKR--EYIMKG	569
Db	585	-YTR-----SMTSEAAAAAATAATEEPPPLPPEPPISNEPETSRRWEEEMEVAKKG	638
QY	570	NINNYFKLGDEGSKYVYHNOYSTNSFALNKQHEDHDKRRHLAKHCLTLPAGEFKWNGS	629
Db	639	LVEHGRNNAALAKMGVTSSECKKNFFNY-----KRRH-----	672
QY	630	VHGSKVLTITLRLTITOLENNIPSSFLHPMASHANWMIKAYOMSKPREFATALAILIE	689
Db	673	-----NLDNLIDQ-----HMOKASR-----KPRE-----EDVISO	697
QY	690	CAVKVYVWLPITREPLGHTLRHRTSTIEREKEKVKKKERKOEETMOATVVKYTPVY	749
Db	698	C-----ESVAST-----VSQOEDIEDIEASNEENPEDSGAENSDE--SABS	739
QY	750	KHOVMKOGEE--YRATGVGCMWSIKTIV-----YFVFKLEGNTVNVRKSL	796
Db	740	PSPVBAASSEDSEMAASRGNTPEVALEATDPAFCASPSAVP-----TTKPAERESV	795
QY	797	ES-----TKNNMDENDESDDRKCS--RSFKIKITIEDSEKDEYKGSADAK--AQ	844
Db	796	EAQVTDASAEATPEAPMDVDHECGAEGSSVLDPPAPYKADSVDPMEQVYENTASKEGGA	855
QY	845	NEMDI--SKITEKKODV-----KELDSDSDKPEER-----PMEV	879
Db	856	KERDEISTSEKTEARDEDDVVAQIERPEPOSDSDSATCSADEGVDEPERQORVPMDA	915
QY	880	DDDKTE-----SH-----VNGQSSQVDVYVNSEGHLRKS	911
Db	916	KPSLITPPGSLISPIKPNLIDLPOLOHAAVAPIPMWSTCPON--PIOTPVSGVAL--	971
QY	912	YKKTKSKSLDGLLERRIKOFTLEEKORLEKILIGIGIGIKGTSTNSKN-----LSES	966
Db	972	YORIKAMHESALLE-----EQORQOQVLDLER--SSTSPCSISKPNMEVLEOPA	1022
QY	967	P--YITAKEGCOSDSMKROSP-----NANDQPEDLIQCSOSDSVLRMSDPS	1015
Db	1023	PHOYITVJPMGCVRLPTTRPRRPPPLIPSSKTTVASEKSPFISGISOGTGTGLYS--S	1080
QY	1016	HTTVKLVKPKXRVLDVVSIRSPEFKCPKXNSIENDIEEKVSDLASGOEPTYSKTKGDNFF	1075

Db	1081	H	NOAYPG	-----	EARKPVSGISGL	-----	PROESTKA	-----	1110																																																																																																																																																																																																																																																																																																																																																																																																				
QY	1076	IDD	SKLASAD	IGTILIC	KNKRP	LI	-----	QOESDITVSSKSA	LHSHVPKSTDRD	ATPLSR	1132																																																																																																																																																																																																																																																																																																																																																																																																		
Db	1111	-----	-----	-----	APLYIKOEERS	-----	-----	-----	-----	-----	1136																																																																																																																																																																																																																																																																																																																																																																																																		
QY	1133	AMD	-----	FEKGLG	CDSESN	STLENS	SDPVS	IQDSEED	MIYON	SNESISQ	FTRODVEY	1139																																																																																																																																																																																																																																																																																																																																																																																																	
Db	1137	AQH	EVG	VGTA	GA	VEG	STIG	TG	PAS	KS	-----	1168																																																																																																																																																																																																																																																																																																																																																																																																	
QY	1190	LEPL	CEL	VSG	ESTG	CED	RLPV	KG	-----	TEANGKRPSQOK	LEERPVNKCSDQ	IKLNTT	1246																																																																																																																																																																																																																																																																																																																																																																																																
Db	1169	ISL	SG	STIG	TPA	-----	LPQAGIT	PE	TAL	VGP	SR	PMPI	EESE	PEK	VEA	SKGHVI	1222																																																																																																																																																																																																																																																																																																																																																																																												
QY	1247	-DKNN	-----	ENNE	SEKK	GORT	-----	SFPO	ING	D	NK	PX	-----	YL	GE	CL	KE	IS	ER	VYSG	1235																																																																																																																																																																																																																																																																																																																																																																																								
Db	1223	YEG	KG	SH	L	ST	D	N	K	A	R	E	G	R	S	P	T	A	H	E	S	L	K	R	S	T	A	E	A	E	V	G	S	I	K	O	M	S	R	E	S	P	A	P	E	G	1282																																																																																																																																																																																																																																																																																																																																																														
QY	1296	NVE	P	V	N	N	I	K	I	P	E	N	D	I	K	S	L	T	V	K	E	S	A	I	R	P	I	N	G	D	I	M	-----	EDFNE	-----	R	1339																																																																																																																																																																																																																																																																																																																																																																								
Db	1283	LI	-----	CRA	L	P	R	G	S	P	H	D	L	K	E	R	V	-----	LSG	I	M	O	T	P	R	A	T	A	E	S	P	E	D	G	L	K	P	K	O	K	R	1331																																																																																																																																																																																																																																																																																																																																																																			
QY	1340	NSE	E	K	S	H	L	L	S	S	A	E	G	N	Y	D	S	E	L	T	P	K	E	S	D	T	O	T	T	P	P	A	S	C	E	N	S	V	N	O	V	E	M	E	T	1399																																																																																																																																																																																																																																																																																																																																																															
Db	1332	ESP	P	R	A	-----	FEG	A	L	T	K	G	P	I	D	G	I	T	I	-----	KEM	-----	G	S	H	E	I	R	O	D	I	L	T	1372																																																																																																																																																																																																																																																																																																																																																																											
QY	1400	SE	V	K	V	-----	T	S	P	I	T	E	E	S	N	L	S	N	D	I	E	N	G	L	P	I	N	K	E	N	V	E	N	G	E	S	K	R	V	T	E	V	T	M	T	S	1454																																																																																																																																																																																																																																																																																																																																																														
Db	1373	QES	R	T	P	E	V	O	S	T	P	I	-----	E	G	S	I	S	-----	O	C	T	P	-----	K	F	D	N	N	S	G	S	A	L	K	H	V	K	S	L	I	G	P	S	1422																																																																																																																																																																																																																																																																																																																																																																
QY	1455	TVA	E	S	K	V	I	K	-----	VE	K	D	K	-----	TV	S	T	E	N	C	A	K	T	V	-----	1486																																																																																																																																																																																																																																																																																																																																																																																			
Db	1423	K	L	P	R	G	M	E	I	V	E	N	I	K	V	E	R	K	E	I	D	V	A	G	E	P	V	A	R	H	T	S	V	S	G	P	S	V	L	R	T	I	H	E	A	R	K	O	L	S	P	1482																																																																																																																																																																																																																																																																																																																																																									
QY	1487	-----	TTTT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1530																																																																																																																																																																																																																																																																																																																																																															
Db	1483	G	L	Y	D	S	S	A	R	T	P	V	O	Y	O	T	I	-----	S	R	G	S	P	M	N	R	T	S	D	V	S	S	K	S	A	S	H	E	K	S	T	L	T	P	O	E	S	I	P	A	S	P	1542																																																																																																																																																																																																																																																																																																																																																								
QY	1531	-----	G	L	V	T	-----	S	M	V	S	K	E	Y	T	R	O	K	V	-----	M	K	S	R	K	T	R	S	T	A	L	-----	1566																																																																																																																																																																																																																																																																																																																																																																												
Db	1543	P	G	V	D	I	V	S	H	S	F	D	P	H	R	S	A	A	G	E	V	Y	S	H	L	P	T	H	L	D	P	A	M	F	H	R	L	D	P	A	A	Y	L	O	R	O	L	S	P	T	P	G	1602																																																																																																																																																																																																																																																																																																																																																								
QY	1567	-P	S	-----	Y	R	K	E	V	T	S	K	E	S	I	F	V	P	N	D	L	K	L	R	K	G	I	R	E	V	P	V	N	K	P	-----	ALD	W	P	V	S	P	R	1621																																																																																																																																																																																																																																																																																																																																																																	
Db	1603	Y	P	S	Q	O	L	A	M	E	N	T	R	O	T	-----	N	O	I	T	S	O	Q	M	Y	N	L	-----	E	D	V	T	R	G	L	S	P	R	O	P	I	G	L	-----	P	V	A	T	R	1656																																																																																																																																																																																																																																																																																																																																																											
QY	1632	P	T	E	G	I	T	R	L	O	V	K	S	L	A	G	V	S	L	M	L	R	L	M	A	S	L	R	W	D	M	A	A	K	V	P	-----	G	G	S	T	R	T	E	S	1674																																																																																																																																																																																																																																																																																																																																																															
Db	1657	G	I	D	L	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	

DR EMBL; AF026245; AAB86384.1; -
DR EMBL; AF083037; AAD22767.1; -
DR EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR EMBL; AF091711; AAD39719.1; -
DR EMBL; AB018346; BAA34523.1; -
DR EMBL; AC000066; AAC60380.1; ALT_FRAME.
DR Genew; HGNC:379; AKAP9.
DR MIM; 604001; -
KW Colled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-RIT SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 914
FT DOMAIN 944 1022
FT DOMAIN 1100 1185
FT DOMAIN 1253 1280
FT DOMAIN 1336 1392
FT DOMAIN 1434 1459
FT DOMAIN 1585 1659
FT DOMAIN 1857 2455
FT DOMAIN 2544 2561
FT DOMAIN 2603 2776
FT DOMAIN 3065 3092
FT DOMAIN 3124 3470
FT DOMAIN 3587 3689
FT DOMAIN 3726 3730
FT DOMAIN 203 292
FT DOMAIN 321 1010
FT DOMAIN 1846 2772
FT VARSPLIC 17 28
FT VARSPLIC 1637 1642
FT VARSPLIC 1643 3911
FT VARSPLIC 2175 2182
FT VARSPLIC 2175 2183
FT VARSPLIC 2895 2907
FT VARSPLIC 2895 2948
FT VARSPLIC 3901 3911
FT VARSPLIC 1347 1347
FT VARIANT 1347 1347
FT CONFLICT 76 76
FT CONFLICT 475 475
FT CONFLICT 554 554
FT CONFLICT 638 638
FT CONFLICT 663 663
FT CONFLICT 913 913
FT CONFLICT 956 956
FT CONFLICT 980 982
FT CONFLICT 997 997
FT CONFLICT 1001 1001
FT CONFLICT 1020 1020
FT CONFLICT 1028 1028
FT CONFLICT 1626 1626
FT CONFLICT 1703 1703
FT CONFLICT 1707 1707
FT CONFLICT 1802 1803
FT CONFLICT 1843 1843
FT CONFLICT 1956 1956
FT CONFLICT 2027 2027
FT CONFLICT 2157 2158
FT CONFLICT 2169 2169
FT CONFLICT 2514 2514
FT CONFLICT 2851 2851
FT CONFLICT 2957 2957
FT CONFLICT 2963 2963
FT CONFLICT 3087 3087
FT CONFLICT 3218 3218
FT CONFLICT 3307 3309
FT CONFLICT 3751 3751
FT CONFLICT 3833 3833
SQ SEQUENCE 3911 AA; 453664 MW; 3FBICBIC819B47AA CRC64;
Query Match 2.7%; Score 380.5; DB 1; Length 3911;
Best Local Similarity 17.2%; Pred. No. 2.7e-06;

Matches 483; Conservative 475; Mismatches 1035; Indels 821; Gaps 115;
QY 305 DCAVET-----OKNP--YIRHEPIGYDRSRKRY-----FLNRRLIE 341
DB 125 DCSSEVNGCSFVARTGKPTMLIREEEGVDDSYEGQAQDSPHLMMESELAGQHEIE 184
QY 342 EDTENENEKKIWIYSTKYVLAELIDCL-----DDYWEALCKTL 381
DB 185 ELNRELEFMRVYTGTEGLOQLOFEPAIKOROGITOLTANIQARREDETWRELELT 244
QY 382 EEMRE-EIH-RHMDITEDLTKARGSNKSFLEANEI-----LESTIRAKKG 426
DB 245 EESOKIQLIOFOQOASETLINSHSTADLLQAKQIILLHQOOLEPDHLEDOYKKKE 304
QY 427 DI-----DNKSEPEIEKKNLENDSKDAEKRE-----EPEDQLEKDS 467
DB 305 DFTMQISFLQEKIR-VYEMEDK-KVENSKEIOEKETIIEBLNTKIIIEEKCKTLELD 362
QY 468 DDKTPD-----DDEQCKSEYGDGFKSEKNGELSEPGAG--KGASGTLITRLRNP 518
DB 363 KLTTADKLGLQEOQVQAKNOELKNNKLELINSKQKROSSSEIKQMGTEVLEQRRNK 422
QY 519 DSKL-----SOLKSQVAAAHBA-NKLEKGEKVLVNSQGEISLSTKEVI 566
DB 423 DSQFETDVIQRMEOFTQRLQELRLAELDEMGOQIVQMKELIRQHMAMQEMDKTRHNGE 482
QY 567 MKNINNYKRLQEGKYRYVHNQYSTNSFALK--HQHEHDKRRHLAKFCLTPAGEF 624
DB 483 MENALSYNI-----TVNEQIKLMVAINELNIKLODTISOKEKEKEEL----- 528
QY 625 KNGSVHSGKVLITSLRLITOLENNIPSPFLHPMASHANWIKAYOMGSKPREPALA 684
DB 529 ---GLLEKCALQRLQELDAVEEL-----SPSRQD--IQRAQIADQ-SKLEAKRS 576
QY 685 LALICAVAPVYMLPIWREFLGHTRLRHTS---IEREKEKVRKKKQOEETMQOA 740
DB 577 LSTVEDLKAELIVASASESRRELELKHAEVYTNKIKLEMLEKKNVLDPRMESAQAEIER 636
QY 741 TWVKTFEPKHHQVMQKGE---EYRVYGGGWSMLSKTHYRVFKLQGNINVRKSL 797
DB 637 LRTQLLFSEELSKLELEIEHRIN-----LEKLDNGLHYKQIID 680
QY 798 GTRKNNDENMD--ESDKRCSRPKIRKIEPDEKDEYKSDAAKAD-----QEM 847
DB 681 GLQNEQSKQIEHQEKONLIRKQMLLEISKLQDOSLVNSKSEEMTQINELQKEI 740
QY 848 DISKITEKK---DDVAKEL-----LDSQDKPCKEEPMEV 879
DB 741 ELTROEKEKGTLEQVQELQKTELLEKQKKEKENDLOERFAQLEAE-NSILDEKCTL 799
QY 880 DDDMKTESHVNQESQYDVVNVSGFHLRTSYKKKTKSSKLDGLLEKRIKQFLLEKQR 939
DB 800 EDMKIHHPVSGEE-----RLIFDSTISKSDSVWEKEI-ELIENED 843
QY 940 LEK--IKLEGGIKIGKSTNSSKNL-----SESPVITKAREGQSDSMQOE---Q 985
DB 844 LKQOCIOLENEIEKQRFNFEPAEKNFENVYQELQEVYCLLKVADLLEDSKNNKELEYKS 903
QY 986 SPNANDPEDIQGCSSDSVSFLMSPSHTKVLPKDRVLDVDSIRSPET----- 1038
DB 904 KIKALINEB-----LHQRINPTVKM--KSSVPEDEKTFVAETLEMEGV 946
QY 1039 -----KCPRONSIENDIEKYSVDLASRQEPKSTKTKNDQFIDSKLASADDI 1087
DB 947 EKOTTELMKELEVTKRELE--LSQRUSDLS---EQKQKHGEISFLNEEVSQKOE-- 998
QY 1088 GTLLCKNNKPLIOESDPIVYSSKSALHSSVPSKSTNDRDAPLSRAMDFGKILGCDSESN 1147
DB 999 -----KEQVSLRCHELEIILHNRA-----ENVSQDITQVS 1029
QY 1148 STLENSDPTYSIQDSEEDMIVONSNEISIQEFTREQDVAVLEPLKCLVSGESTGNCE 1207
DB 1030 STL---DGVYTMNTRGAEGSVKNSFGEESEKINVEDKVSFE---NMTVGEESKQEQ 1081

QY	1208	DRLP	-	VKGTANCKAKRSOQKLEBRPVAKCSOQIKATITDKNNENRESEKQOR	1262				
Db	1082	LILHLPSVKESSLRATQOSENDKQ	-	KELVNLKSEONDLR-LQMEQOR	1129				
QY	1263	-----	TSFOI	-----	NGKNNKRIYLKGCJCLKEISESNVSGNPEVYNNINKIIP	1309			
Db	1130	ICLSLVYSTHVDQVREHMEKQ	-----	KALCSLKEELIFA	-	OBEKIKELQK-TH	1177		
QY	1310	ENDIKSLTVKESALRPINCQ	-----	VIMEDFENENSETKSHLLSSDAEGNYRSL	1363				
Db	1178	OLEJOTKMOFT	-----	GDECKPLHLILGKLOKAVSECGYFIOTICSVLGEY	-----	1225			
QY	1364	TLPSKCSDSQTITPASCPEENSVQVEDMIEHSEVKKYTSPIITSEESNSNDFI	1423						
Db	1226	-----	YTPLAK	-----	EVANDEKEN-SGYI	1246			
QY	1424	DENGLPINKN	-	ENNVAGESKRTVITEVY	-----	TMTSTVARESKTVIKVEKQD	1470		
Db	1247	SENEDPELOJRYREVQOFQENMHTLNNKVTAEYNKLVLQTRLSKIMQOQDGMKLEGE	1306						
QY	1471	KQYVSSSTENAKSTVTTTTYTKLSTSTGSDVDIISVKQSKTVYTTVTDSLTGG	1530						
Db	1307	-----	ENLPKE	-----	ETEFLSHSQ	-----	1322		
QY	1531	GLTVSMTSVSKSEYSTRDKVILKMFSPRKRTRSGTALPSYRKRVTKSKJFVLPNDDLK	1590						
Db	1323	-----	MTNLEDIVNNKSKLSSLODELTKLEQVQELISLSSLOQO	-----	LK	1367			
QY	1551	KLARKGGRREVYFPFNNAKALDIMPYPSBPRTFGITWRYRLQTVKSLAGVSLMLRLMA	1650						
Db	1368	ETEQ	-----	NYEAE	-----	THCQO-KRLQVANS	1388		
QY	1651	SLRMDMAAKVPCGSGSTRJETSSETLTTTEILKRDRQVPGIREVCIRKICQIGPE	1710						
Db	1389	-----	ESTYPP	-----	SLPVDVSVITTESDAQTMFGS	-----	CVKNNI-DGITE	1427	
QY	1711	TPKETPPQKRGKSSALRPRKREPTKQTPVILITETWA	-	EBEELMWEITRAFAEREKEX	1769				
Db	1428	FSGEGFVKEETINIVKLEKQYOQOLEEBAKAVIVSMAIFAQOTELSRISGKEHTASK	1487						
QY	1770	-	AOAVEQAAKRLQEOQPYIAIISTSPSSSTRTISPAQKMAVAPISGVTYTGKMYLT	1828					
Db	1488	QAHVAQCOBOHYENEMK	-----	LSQOIGQFFETVDVAKFKEEPRLSKEJGHEKRELL	1542				
QY	1829	TKVSPATVTFQOQKNFHQTFATVWQOGSNSGVVOQOKVGLIIPSSSTGSOQFTSFQ	1888						
Db	1543	S	-----	NSDPHD	-----	IPESKNCVILT	-----	SEMFESDK	1569
QY	1889	PRATVATIRPMTSSGSGTTSNSOYITGPQIRPMYVIRPILQOOSTLGAIIIRTPVMOVQ	1948						
Db	1570	TFIVQROSIHETISVSSMDARQMLNEEDLEDMROELVQYQEOQATELLROAHROME	1629						
QY	1949	APQOVMQIIRGQPVSTAVSAPMTVSSSTGQKSLTSATSTNSIQSASQPPROQOQVLT	2008						
Db	1630	RQREDOQL	-----	QBEIKRLMRQLOAQSISIDNELVSERRVILBELLELKLKOLSLAGREKL	1666				
QY	2009	TMAQLTQITQGHGNOGLIYVIGOGQTTGQ	-----	LQILPQGV	-	TVLPEQOQLOMQA	2060		
Db	1687	CCELRRNSTQTONQ	-----	ENQGEVEEQFKEKEJLDRKEDVPPILISNERVALOKA	1739				
QY	2061	AMPNGATVQREFLPPLATTATTA	-----	TTTTVSTTAAGCBGROSKLSQMOVH	2111				
Db	1740	-----	NNRLKILLEVYKTTAAVETIGRHVLGLIDRSSKQSSASLIMSEAEVAKSCVH	1766					
QY	2112	QDKTLP	-----	PAQSSSVGP	-----	AKAPOQTAQPSARPQOPQOSPAPQ	-----	2153	
Db	1797	EEHRTVDESIPSYSGSDMPBRNDINMMSKVTEBCEGELQSLRVRSQFATETIDPENBELML	1856						
QY	2154	-	VQIQPEVQOTYTVSSNVSEAOPTAAQSSKQVAAQS	-----	QPOSVAGQSPVRYOSP	2207			
Db	1857	NISSRLQAAVEKLEELASSETSSOLEHAKYTQETLIMKRESPROKQEZETBESLKOCEELIREHL	1916						

QY	2208	SQTRRRSTPQLSPGQ-----QSQVQTT-----SQPIQPHSTLSIQPS	2244
Db	1917	EESSRAREDLAAVELSEKAGVIGVADENKTEFLERRIOEKTDIDIRLEQELLCAASNRLQDELTA	1976
QY	2249	QGGPQSQPQVQVSSSTQTLSSGGTLNQVSSVSPSPRQIQIQPPQVIAVPQLQQVQVQVLSQ	2306
Db	1977	E-----QQQIDIEBELLRS-----QKEMKADAG-----VED-----QLIQETFKLKK	2015
QY	2309	IQSQVVAQIQQAQSGVPPQIK-LDLPPIQIQSSAVQTHQIQNY-----VTVQAASVQEQQLQ	2363
Db	2016	EKLVEVQQAQAEKVRBDLDQKVALEIDVDEQVSNFLEQEKNTLMDLQQAQNALEKQLE	2075
QY	2364	RVQQLRDPQQAQKQKQQIQEIKREHTLQASNOSEITQK-QVVMKHA-----VIEHLKQK- 2411	
Db	2076	KMRFLLEQALIDREHERDVQEQEIQLEQDLKVVPFPQPISEHQETREVEQLNNHLKEKD	2135
QY	2417	-----SMTPRAREENQMIYCNQVMKTIIDKID-----KESKQA-----AKKKEESVE	2465
Db	2136	KCSBELLSKEDQLQNDI-----QERNNEIEKLEFLRYRELEQALNYSADTFQKVEQRKHPGAVE	2199
QY	2462	QK-----RSKQATKLSALFLFKHKEHQAELTKRALLDK-DIQIEVQ	2503
Db	2193	AKPELSTLEVOQLAARDAIDRKREKITLLEBQL-----EQPRELEKNNNEVQQLHMQLEIQ	2246
QY	2504	-----EELKRDLTIKKEKEDLMQLAQATVAPARCPRTVYLAPRABPSPPPPPGVQ	2555
Db	2249	KKESTTRLOEQLQENKLFKK-DMEKTLGLA-----IK	2276
QY	2556	HTGLSTPPLTPVPAQK-----RKREEKD-----SSSKSKKKMISTSTSEKTKKDKTKLYCI	2604
Db	2279	ESDASTQDQVHLEFGKFAQIIQEEVEIDQLNEQVTKRLQDQLKITTDNKEVIEKKMELLRD	2335
QY	2607	CKTPYDESKFPIGCDRCQNNYHGRCVGIIDQSEAEILIDEVYCPQCSQSTEDAMTVLPTLEK	2666
Db	2339	LET-----QIEC-----LMSQDECK-----RRREIEQLQNLMEVIEK	2376
QY	2667	DYEGIKRV--LRSLQAHKMMAPPLEPYDPNDADYGVIKREPMDLATMEERVQRRYREKL	2724
Db	2371	LQOELANIGQKTSNNHSLTS-----EADS-----LKHQDVIYIAEKLALEQOVETA	2417
QY	2725	TEFVADMKIKIFDNCRYYNPSPDSPPYQCAEVLSEFPOKLGKFGASNNHKIQS	2778
Db	2418	NEEMTFMKNVKLKTFN-----KMNQLTQELFSLKRESEVEKIQS	2457
RESULT 8			
PHP_DROME			
AC	P39769;	STANDARD;	PRT: 1589 AA.
DT	01-FEB-1995 (Rel. 31, Created)		
DT	01-FEB-1995 (Rel. 31, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Polyhomeotic-proximal chromatin protein.		
GN	PH-P.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Imaginal disk;		
RX	MEDLINE=92146957; PubMed=1346609;		
RA	Decamillis M., Cheng N.S., Pierre D., Brock H.W.;		
RT	"The polyhomeotic gene of Drosophila encodes a chromatin protein that		
RT	shares polyene chromosome-binding sites with Polycomb.";		
RL	Genes Dev. 6:223-232(1992).		
RN	[2]		
RP	SEQUENCE OF 1991-1584 FROM N.A.		
RC	MEDLINE=92039031; PubMed=1937015;		
RA	Deatrick J., Daly M., Randsholt N.B., Brock H.W.;		
RT	"The complex genetic locus polyhomeotic in Drosophila melanogaster		
RT	potentially encodes two homologous zinc-finger proteins.";		

RL Gene 105:185-195(1991).
 CC -1- FUNCTION: BINDS TO POLYTENE CHROMOSOMES. SEEMS TO INTERACT WITH
 CC PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER
 CC COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND
 CC SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF
 CC OTHER PAIR-RULE GENES SUCH AS EVE, FTZ, AND H.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: SALIVARY GLANDS.
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
 CC -1- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X63672; CAA45211.1; -
 CC EMBL; M64750; -; NOT_ANNOTATED_CDS.
 CC PIR; S23632; S23632.
 CC FLYBASE; FBgn0004861; ph-p.
 CC InterPro; IPR001660; SAM.
 CC Pfam; PF00536; SAM; 1.
 CC SMART; SM00454; SAM; 1.
 CC PROSITE; PS50105; SAM_DOMAIN; 1.
 CC Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
 FT ZN_FING 1365 1387 C4-type.
 FT DOMAIN 1513 1577 SAM.
 FT DOMAIN 74 80 POLY-GLN.
 FT DOMAIN 411 450 GLN-RICH.
 FT DOMAIN 494 520 GLN-RICH.
 FT DOMAIN 619 650 GLN-RICH.
 FT DOMAIN 775 960 GLN-RICH.
 FT DOMAIN 1233 1290 SER/THR-RICH.
 FT CONFLICT 254 254 MISSING (IN REF. 2).
 FT CONFLICT 1415 1415 D -> A (IN REF. 2).
 FT SEQUENCE 1589 AA; 167297 MW; A6DF0CF9106E1891 CRC64;
 Query Match 2.6%; Score 379.5; DB 1; Length 1589;
 Best Local Similarity 21.3%; Pred. No. 1.1e-06;
 Matches 242; Conservative 137; Mismatches 407; Indels 349; Gaps 42;
 QY 1632 LQYKSLAGVSLMRLIMASLRMDMAKVPG-----GGSTRLETSETETTTETITIKR 1685
 Db 93 LETLAQKAGITPEKY-----DVASPPHPIAQOQATSGTPATGSGSVPTGS--HR 142
 QY 1686 RDVCPYGIREFYCRKILICPIGVETPKETPTQKGLBSALRPKRPETPKQTGPYIIE 1745
 Db 143 HGPPPTGR-----QHTTP-----STPNRPSPSTPNINCNSI-- 175
 QY 1746 TWVAEELLELMEIRAFERKEKAQAVEQAKKRLKQKPTVIATSTSP----- 1797
 Db 176 -----ARTSLTLKKAQNGQ-----VAATTVPLOLISEQOL 209
 QY 1798 -----SSTTSTSPQKYNVADISGVTTGKMLVTVKVSFATVFOQNKNFHOTFATWK 1854
 Db 210 QFYASNPYALQVKEPPTHTTSGS---GTELKHAQNI---MEVQOQQLQOQLEANG 261
 QY 1855 QGQSGNSGVVQVQKVLGIIIPSSITGSOQFTTSQPTATVITRP-----NTSGSGGTTSN 1909
 Db 262 GGAASAGAGGAAS-----PANSQSOQO---QHSTAISTWSPMOLAATGAGVGDWTVQ 311
 QY 1910 SQ-----VITGPQIRPGMTVITRPLOQSTGLKALLRTVVMQPGAPQV 1953
 Db 312 GRVQLMOPSTSFPLQPMIVSGMLLHPG-GLQOQPIQVITACKPF-----QGNCPQML 363
 QY 1954 MT-----QIRGQ-----PVSTAVSAPNTVSSSTPGKSLSATST 1988
 Db 364 TTTTQNAKOMIGQAGFAGNATATCIPTNHNSPQTVLFSPPNVNLSPPQOQNLDS---- 419

QY 1989 SNIOSSASQPPRRQGGYKYLMAQTLQITGCHGNGCLTVVIGQG-QTTGQL-QLIPQG 2046
 Db 420 ---MAAAQOQOQLQOQOQOQFNQOQOQOQOQLO---QOQOQITLALAVGVDAQKGLQKLVQK 473
 QY 2047 VTVLPGPQOQOQAMPNGT-----VQREFLPLATATTATSTTTVYSTTAAGTGE 2098
 Db 474 VTTT-----SSAVQQAATGGSTGSGTQTOQVQVQVQOQOQOQTQTTQOCQVQSVSTLIPVGVGG 529
 QY 2099 QROSKLSPQMOVHODKTLPPAQSSSVGPBAQ--PQTAQSPARQP-----QTPP-- 2146
 Db 530 Q-----SVQTNQLNLAGOAOQOQMLPWFQLNNAAGLQPGFPQITLRNPDG 574
 QY 2147 -----QSPAQPVQTOPE--VQ-----TQTTVSSHVSEAQPHQAQSKFQVAAQSQPOS 2194
 Db 575 TQGFHTQOQRPQVTLQTOQNOQITQCNVQTPYKARQLDLALAPKQOQOQOQVQVTTQTOQ 634
 QY 2195 NVQGSQFVRVQSPSQRTI-----RPSTP-----SOLSPQGSQOVQTTTSQPI----- 2236
 Db 635 QQLAVATAQIQOQOQOQLTAAALQRPGARVPMRHNQTVRPAQSVSTQTAQNSILKAKMRN 694
 QY 2237 ---PIQPHSLQIPIBSGQPOSQPOV-----QSSTP----- 2263
 Db 695 KOQFVRPALATLKTETIGQVAGQNKVGHLLTVQOQOQATMLQOVNNAQNKVVMSTTGT 754
 QY 2264 --TLSSGQTLNQVSSPSRPOLOIQRPQVYAVRQLOQOQVYLSQIQSQVVAQIQAOQ 2321
 Db 755 PTLQNSQTLHAATAMAVDQOQOQLQFOKQOI--LQOQ-----QMLQOQITAAIQMOQ 805
 QY 2322 SGVPQOQIKLQPIQIOSSAVQTHQIQNVVTVQAASVQEQLOQVQVQLRQOQKQKQOIE 2381
 Db 806 Q-----QAAVQAOQOQOQVQSQOQVNAQOQVNAQOQVNAQOQOQVQOQOQVNAQ 857
 QY 2382 IKREHTLQASQSEIIOQOVVMKNAVIELKQKKSPTAEREQRMIVQNCVMKYIID 2441
 Db 858 AQAOHQOALAMATQOILQ--VAPNQETTS-----QOQOQOQOQILHQDLOQOQLQ 903
 QY 2442 KIDKEKQAKKRRKEESVQKRSQKQATGSLALFKHKEQLRAELTKKRALDQLOIE 2501
 Db 904 QQAQAVQV 936
 QY 2502 VQELKRDQKIKKQKMDLQOQTAQVAAQPCPVTPVLA----- 2540
 Db 937 VQSGATSOQTSQOQOQHNSGQLOLSSVPFVSSSTTPAGIATSSALQAALSAGAIQGT 996
 QY 2541 --PPAPPPPPPPPPGVQHTGLSTPLPVASQKRRKEEKSSSKKKMMISTT 2593
 Db 997 AKGTGSSSSPSTSSVVTITNOSTPLVTSSIVASIQOQQTQSAQVNHQOQLISAT 1051
 RESULT 9
 ANK2_HUMAN
 ID ANK2_HUMAN STANDARD; PRT: 3924 AA.
 AC 001484; 001485;
 DT 01-APR-1993 (rel. 25, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, noneye/thnoid).
 GN ANK2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Brain stem;
 RX MEDLINE=91302466; PubMed=1830053;
 RA Otto E., Kumamoto M., McLaughlin T., Bennett V.;
 RT "Isolation and characterization of cDNAs encoding human brain
 RL j. Cell Biol. 114:241-253(1991).
 RN [2]
 RN REVISIONS.
 RA Carpenter S.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE-Brain stem;
 RX MEDLINE=94075409; PubMed=8253844;
 RA Chan W., Kordeli E., Bennett V.;
 RT "440-kD ankyrinB: structure of the major developmentally regulated
 RT domain and selective localization in unmyelinated axons."
 RL J. Cell Biol. 123:1463-1473(1993).
 RN [4]
 RP SEQUENCE OF 463-495 FROM N.A.
 RX MEDLINE=92009921; PubMed=1833308;
 RA Tse W.T., Meminger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
 RA Lux S.E., Ward D.C., Forget B.G.;
 RT "Isolation and chromosomal localization of a novel nonerythroid
 RT ankyrin gene."
 RL Genomics 10:858-866(1991).
 CC -1- FUNCTION: Attach integral membrane proteins to cytoskeletal
 CC elements. Also bind to cytoskeletal proteins.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
 CC CELLS THROUGHOUT THE BRAIN.
 CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
 CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
 CC AND FUNCTION (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 CC EMBL: X56957; CAA40278.1; -
 CC EMBL: X56958; CAA40279.2; -
 CC EMBL: 226634; CAA42644.1; -
 CC EMBL: M37123; AAA62828.1; -
 CC PIR: S14533; S14533.
 CC PIR: A39643; A39643.
 CC PIR: B39643; B39643.
 CC PIR: S14569; S14569.
 CC HSP: P42771; IDC2.
 CC Genew: HGNC:493; ANK2.
 CC MIM: 106410; -
 CC
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000488; Death.
 CC InterPro: IPR000906; ZUS.
 CC Pfam: PF00023; ank; 24.
 CC Pfam: PF00531; death; 1.
 CC Pfam: PF00791; ZUS; 1.
 CC PRINTS: PRO1415; ANKYRIN.
 CC SMART: SM00248; ANK; 21.
 CC SMART: SM00005; DEATH; 1.
 CC DR PROSITE: PSS00218; ZUS; 1.
 CC DR PROSITE: PSS0088; ANK_REPEAT; 20.
 CC DR PROSITE: PSS0297; ANK_REPEAT; 1.
 CC DR PROSITE: PSS0017; DEATH_DOMAIN; 1.
 CC DR PROSITE: PSS0017; DEATH_DOMAIN; 1.
 CC KMW Cytoskeleton: Alternative splicing; Repeat; ANK repeat;
 CC phosphorylation.
 CC
 CC FT REPEAT 63 92 ANK 1.
 CC FT REPEAT 96 125 ANK 2.
 CC FT REPEAT 129 158 ANK 3.
 CC FT REPEAT 162 191 ANK 4.
 CC FT REPEAT 193 220 ANK 5.
 CC FT REPEAT 232 261 ANK 6.
 CC FT REPEAT 265 294 ANK 7.
 CC FT REPEAT 298 327 ANK 8.
 CC FT REPEAT 331 360 ANK 9.
 CC FT REPEAT 364 393 ANK 10.

FT REPEAT 397 426 ANK 11.
 FT REPEAT 430 459 ANK 12.
 FT REPEAT 463 492 ANK 13.
 FT REPEAT 496 525 ANK 14.
 FT REPEAT 529 558 ANK 15.
 FT REPEAT 562 591 ANK 16.
 FT REPEAT 595 624 ANK 17.
 FT REPEAT 628 657 ANK 18.
 FT REPEAT 661 690 ANK 19.
 FT REPEAT 723 756 ANK 20.
 FT REPEAT 727 756 ANK 21.
 FT REPEAT 760 789 ANK 22.
 FT REPEAT 793 822 ANK 23.
 FT DOMAIN 1773 1950 REPEAT-RICH REGION.
 FT REPEAT 1773 1784 REPEAT A.
 FT REPEAT 1785 1796 REPEAT A.
 FT REPEAT 1797 1808 REPEAT A.
 FT REPEAT 1809 1820 REPEAT A.
 FT REPEAT 1821 1832 REPEAT A.
 FT REPEAT 1833 1844 REPEAT A.
 FT REPEAT 1845 1856 REPEAT A.
 FT REPEAT 1857 1867 REPEAT A.
 FT REPEAT 1868 1879 REPEAT A.
 FT REPEAT 1880 1891 REPEAT A.
 FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
 FT REPEAT 1903 1914 REPEAT A.
 FT REPEAT 1915 1926 REPEAT A.
 FT REPEAT 1927 1938 REPEAT A.
 FT REPEAT 1939 1950 REPEAT A.
 FT DOMAIN 3536 3620 DEATH.
 FT VARSLIC 1039 1039 Q -> QFGKHLPTAPPLINRGESLVSRIIDGPRGK
 FT VARSLIC 1444 3528 (IN ISOFORM 2).
 FT CONFLICT 475 476 MISSING (IN ISOFORM 2 AND ISOFORM 3).
 FT CONFLICT 971 971 GO -> PE (IN REF. 4).
 FT CONFLICT 971 971 I -> S (IN REF. 1).
 FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).
 FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
 SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;
 Query Match 2.6%; Score 375.5; DB 1; Length 3924;
 Best Local Similarity 17.7%; Pred. No. 4,1e-06;
 Matches 550; Conservative 398; Mismatches 1121; Indels 1037; Gaps 126;
 QY 16 TDSEDEDEDEDDDDSDYPREMEDDDDASYCTESSFRSHSYSTSPGRKRPVHRP 75
 DB 1423 TKSESDQEQEEFID-----MTSEKNDETESTETSLVLSHLVNEVVLASPLD--- 1470
 QY 76 RPILEEK-----DIPPLEFPKSSSEDLAMPNEHINAVIAYELRN 116
 DB 1471 LSEVSEKKODLIRMTAILTTVDSDKAGSIRKELVNAABE--BGEFFELVERKEDLEK 1528
 QY 117 FGVVLRSPREFDFCAALVSOCETLMAEMHVLLKAVIREDDTSNTPGPAUL----- 171
 DB 1529 VNEILR-SCICTDESSVSSRSERGLVEEWYI-----VSEDEIEAROKAPLEITIEYPC 1583
 QY 172 -----KDSVNSTLYFIDGMW-----PEVLRYVCESD 198
 DB 1584 VEYRIDKEIKGVKEDSTGLVNLITLDDNLCVLPKQLOLYODKAGKKKEALAVGSSSE 1643
 QY 199 KEYNHVP-----YQEAEDYRG---PVENKIKVLOFLVDQFLTNTIAREELMSEGVLYQ 250
 DB 1644 KEKKDIPRDETOSTOKHNPISLGIKRPVRRKLEKQKQKEEGLOASAEKAL----- 1695
 QY 251 DDHCGRVCHKGLDILCCETGSAVYHNLGCVKRPYLEVPEDEMCVCVAHKVPGVT--DCVAE 309
 DB 1696 -----KKGS-----SESLGSD-----PGLAPPLPT 1717
 QY 310 IQKNKYRIHEPIGDRSRKRWFLNRLIIEEDTENENKRIWYSTVKQALIDICD 369
 DB 1718 VKNTSPLIETPIGSIKDKVK--ALQKV-----EDQK----- 1749
 QY 370 KQYWEALCKILEMRBEIHRHMDITEDLTNKRAGSNKSFSLAANBELISIRAKGDID 429

Db 1750 -----GRSKLPIRVKCKEDVPKKT--THRHPHAPSS--LKSERHAPG--- 1788
 QY 430 NKSPEETKDNKTENDSNDAEKNREFEEDQSLKOS----DDKTPDDDEQCKSEVGD 485
 Db 1789 ---SP-SKTERHSTLSSAKTERHPVSPSSKTEKHSVPVSAKTERHSPASSSS---- 1840
 QY 486 FKSEKSNELSESPACGAGSSTRILIRLNNPDKLSQLKSOQVAAAHANKLFKEGK 545
 Db 1841 -KTEK-----HSP-----VSPSTK--TERHSVSTTEHHPVSPSG-----KTDK 1879
 QY 546 EVLVNNSOGELISRLSTKEVIMKGNINNYFVLGQDGKRYVHNQYSTNSPALNKHQRED 605
 Db 1880 RPPV-----SPSGRTEKHPPVSPGTERKRLPVPSPSG-----TDKHQPVST 1920
 QY 606 HDK-RRHLAHKFLTPAGEFFKNGSVHOSKVLIJTLRLITTOLENNIPSEFLHPNMASH 664
 Db 1921 AGKTEKHLP-----VSPSKTEKOPVSP--TSKTERIETTMKVREL----- 1960
 QY 665 RANNIKAVOMCSKPREPALALILECAVRYMLPIWREFLGHTRLHRMTSIEREEKYV 724
 Db 1961 ---MKAFOSGQDPKSKHKTGLFEKHSKQK-----QOQEKGV 1994
 QY 725 K-KKEK-----KOEFEETMOATWVKTTPYKHOVMKQGEVRYVGYGKMSLTKTHV 778
 Db 1995 RVEKEKPIILTOREAQKTENQTI-----KRQORLEPVTG----- 2027
 QY 779 RFVFKLPGNTVWYRKSLGKTKNNMDENDESDKRCRSRPRKIKIEPDEKDEVKGSDA 838
 Db 2028 -----TAESKRGVRSSTIGVKK--EDAAGKEKVLSHKRIPEVQASVPEHSEHSEVPR 2079
 QY 839 AKGAD-QNEMDISKITEKK-DQYKELLDSDSPCKEPEMEVDDDM-KTESHVNCOESS 895
 Db 2080 EKMADGEGDDMLQISPRKSTSESEVIAKQELDENKQOQFRLSEETEKQALHLD----- 2134
 QY 896 QVDVYVNSSEGFHLTYSYK-----KTKSSKLDGLE-----RRTKOP 932
 Db 2135 QVLSPNTTFPL--DYMKDEFLPALSLQSGALDGSSESLKNGVAGSPGCSLMEGTPOI 2192
 QY 933 TLEBKQLEKIKLEGKIGKTSNNSKNLSBPVITRAKEGOSDSMRQEOSPANND 992
 Db 2193 SSESSEYKHE-----GLAETPETSPELSFSPKSEQGTETKSTFTTTTTERSE 2243
 QY 993 Q-----PRDLQOGSOSSVSLRKMSDPSHTNKLKPKDRVLDVYSINSPETCKCKONSTEN 1048
 Db 2244 KEHPHTKIDITGSEBERQAV--TEDESETSESEFOKEATLG-----SPKQTSPRK--QD 2292
 QY 1049 D-----IEEKVSDLASR--GOEPTKSKKGNDFIDOSKLASADDIGI 1089
 Db 2293 DCGSGSVALAKETPIGLTEAACDGOQRTFGSSAHKTOT-----DSEAOES-----T 2340
 QY 1090 LITKNNKPLIOESDITVSSSKSALHSVPRSTNDRDAPPLSRAMDFEGKLGCDSESNST 1149
 Db 2341 ATSDETKALPLPEASVKTDTGTGTESKPOGVIRSPQGLELALPSR-----DSEVL5A 2390
 QY 1150 LENSSTPVSIOGSEEDMIVONNESISBQFTRREDVAVLEPLKELYSGESTGCEDR 1209
 Db 2391 VADDSILAVSHKDLSEASPVLEDNSS-----HKTPDSELP-----SPKESPCRS 2435
 QY 1210 L---PVKGTSEANGKPS-----OQKLEERPNV 1234
 Db 2436 LESSPVEPKMKAGIFPSHPPLPAVAKTELLETVASVRSRLRDPGSMADSDLEGTSLM 2495
 QY 1235 KCSDOJLKLANTYDKNNENRESEKKGQORTFOJINGKDNKPKIYKLGCEKLKTESRYVS 1294
 Db 2496 ESSGKSPILSPDPPSSSEVSEYVTPKTTDVST-----PKPAVTHECAEEDD-----S 2541
 QY 1295 GANEPPVNNINKIIPENDIKSLTVKESAIRPFLINGDVIMEDNENRNSSTKSHLSSDA 1354
 Db 2542 ENEKER-----RTPPEEFKMYTKLTKMDELQEAQKQKRDYKKEPKQESS---SSDP 2593
 QY 1355 EGNYRNDLETLPLSTKSDSTQTTTPSASCPESSNSVQVDEMELETSEVKKYNS----- 1407
 Db 2594 DADCSVDVDE-PKHNGSGDESGVPLVYHSEKRYSSSESESEPELQALKKAGDSGLLPP 2652

QY 1408 -----SPITSEESNLSNDFI-----DENGPIPNKNENNGES----- 1440
 Db 2653 VIRQVPPSPILSSNDSNNSPEEVOFQPVVSKOYTFKKNEDTOEERPGKSEBEKSESHLAE 2712
 QY 1441 KRKTIVTEVMTSTVATESEKTVIAKVEGDKQVYVSSTEKCASTVTTTTTYTKLSTPS 1500
 Db 2713 DRHAVSTE-----AEDRSYDKLNRDIDQKICDHGCEAMSPSSARPPVSSGIQSP 2763
 QY 1501 TGGSVDIITSVEKOS--KTVVTTTTVSDSLTTTGGTLVTSMT-----VSKESTRDKV 1550
 Db 2764 TGDVD-----EQVYIKESIALQGTHEKDTQEGEILDVNSAESBPQADCPSESSSSSLP 2818
 QY 1551 LMKFSRPRKTRSG-TALPSRKEVYTKSTKSTIVLPBDDLKTLARKGGLREVPYFNWNAK 1609
 Db 2819 HCLVSEKKEDEDEISATSSIOKTEVTKTDETFENLPKD----- 2856
 QY 1610 PALDIWYPPSRPPTFGITWYRLOTYKSLAGVSJMLRLMASLRMDMAKVPBGGSSTR 1669
 Db 2857 -----CPSQDSST--TOT-----DRFSMDVVSOLAEN 2883
 QY 1670 TETSETEITTTETIKRQDVGPVGIREFYCIKILCPGVP-----ETPKETPPOR 1721
 Db 2884 DEIYDQITS-----PYE-----NVPQSFFSSESKTQTDANHTT 2919
 QY 1722 GLRSSALRPRKRETP-----KOTGPVITETWAEHELELEIRAFARVEKEKAQAV 1773
 Db 2920 SFHSSEYVSTTSPVEDVYVASSSGTVLSK---ESNFEQDGIK-----M 2962
 QY 1774 EQQAKRRL-EQKPRVYATSTSTPSSTSTISPAQKVMVAPISGVTGTGKVLV-TRY 1831
 Db 2963 ESQLESTLMEWQSDSV--SSSEFPTMSATTTVGEQI-----SKVITIKTV 3007
 QY 1832 GSPATYTFQONKNEHOTFATWVQOGSNGSVQVOQKVLGIIPBSGTISQOFTSFQPT 1891
 Db 3008 DSD--SWEIRDEDAFEARVK-----EEBKIRGLM-----YDRQ----- 3041
 QY 1892 ATVTIRPNTSGGTTNSQVITGPIRPGMTVIRPPLQOSTLGNKILITPMVQPGAPQ 1951
 Db 3042 -----SGIT-----PDITPARIPTTEGT--PRSEONPFLPBGK-- 3074
 QY 1952 QVMTQIIRQOPVSTAVASAPNTVS--STPGOKSLTATSTSNIOSASQPPRPOGQVYL 2008
 Db 3075 --LEFMTRSGALIDMTKRSYADESFHFQIGQESREBTLSYEDVKEGATGADPLPLE----- 3127
 QY 2009 TMAQLTQLQGHGNGGLTVIIOGQQTGLOLQIPQVT-----VLPGQOQLMOAMPN 2064
 Db 3128 TSAESIALS-----ESKETVDEADLLPDSVSEVEEELIPASDAQLN----- 3168
 QY 2065 GYVQRFLEPLPLATTATTATTTTSTTAAGTGEORQSKLSPQMVQVHODKTLPPAQSSSV 2124
 Db 3169 -----SOMGISASTETPTKEAVSVGI-----KDLPTVQTDGI 3200
 QY 2125 GPKAQOPQTAQ-SARPOPOQ-----POSBAQ-----EVQTOBEV 2160
 Db 3201 PPLSGVKQIQCPSDSEPAVOQLDFSTLRSVYSDRGDSDPSSEBQKSVIEIPTAME 3260
 QY 2161 QTOQTWVS-SHVSEAOPT-----HAOSKSP----- 2184
 Db 3261 NVPFTESKSKITVRIIMPSTPAIPSAEYESSYSEDFLSSVDEENKADBAKPSKLPVKVP 3320
 QY 2185 -QVAAOSQOSNVQOSQPVVQSPQTRIRPSTPQSLSPGOOSQVQTTSPPIQOPTS 2243
 Db 3321 IQRVQOQLSDLDLTVQKVIAPQGOUMASIAPDNRK--SESDASSLDSKTKCPKVTKRSTTE 3379
 QY 2244 LOIPQGOQO-----SOPVQOSTQTLSSQTLNV--SVSPSRPOLQ----- 2285
 Db 3380 TETESRERAELELESEEGATRPKILTSRLPYKSRSTISSCRGTSPTKESKEHFDLYR 3439
 QY 2286 -----IQQPOQVYLAIVPQLOQOVVLSQIQSQVVAQIQAOQSSVPPQIKIQLPQI 2336
 Db 3440 NSIEFFEISDASKLVRLTQOSEREQIVSDDESSALAEVSVIENLPVETEHSEVPEDI 3499

DR PDB; 1VOM; 23-DEC-96.
 DR PDB; 1LVK; 28-JAN-98.
 DR DictyDb; DD01008; mhca.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAD.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS0096; IQ; 1.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
 FT DOMAIN 762 791 IQ.
 FT NP_BIND 179 2116 COILED COIL (POTENTIAL).
 FT DOMAIN 638 660 ATP.
 FT DOMAIN 738 752 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
 FT MOD_RES 678 678 ALKYLATION (SH-1).
 FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770B1E56A1 CRC64;

Query Match 2.6%; Score 374.5; DB 1; Length 2116;
 Best Local Similarity 19.6%; Pred. No. 2.3e-06;
 Matches 347; Conservative 275; Mismatches 626; Indels 523; Gaps 78;

QY 31 DDDSDYPEEMEDDDDDASYCTESSF--RSHSTYSGTGRKRPVHRPRSPILE----- 81
 DB 530 DEOSVFPN-----ATDNTLTITKLHSHS-----KKNKAYEPRRSKTEFGVTHA 574
 QY 82 -----EKDIPPLE-----FKSSEDLAVNEHIMANVIAIYELNFGVLRIS 124
 DB 575 GOVYVIOIOWLEKCPLOODELELCFKDSSDVNVTKLFNDPNIASRAKGANFIV----- 630
 QY 125 PFEFEDFCALVSEOCQTLMAEMHVLLKAVLREEDTSNTEFG-----PADLKMS 174
 DB 631 -----AAQYKEQLASLMAFL-----ETTPHVFRCIIPNNKOLPAKLECK 670
 QY 175 VNSTVFIDGWTPEVLR-----VYCESDKEYHHVLPY--OEAD----- 212
 DB 671 V-----VLDQLKNGVLEIGIRITRKGFPRRIIYADFVKYIYLAPVPRDADDSQKATDA 725
 QY 213 -----YPYGVENKIKVLOFLVDQFLTTINAREELMSEGV--IQYDHCRCVCHK 259
 DB 726 VLKHLNIDEOYRFG-----ITKIFFRAGQLARIEAREORISSEIIKAIQOATRGMIARK 780
 QY 260 LSDLCCETCSAVYHLECYKPPLEVEPEDEMOCCEVCAHKVGVNDCAVALOKN-KPYIR 318
 DB 781 -----VY-----KQAREHTVAARI-----IQDNRAVI- 803
 QY 319 HEPIGDGRSRRYW-----FLNRRLIEEDTENENKKIWMYSTFYQVLAELIDCDK 370
 DB 804 -----DFKSWPMKLFESKARPLLRNPEKEIKERE-----ILEKSNLTD--STQK 851
 QY 371 DYWEAELCKILEEMREELHRRHDTEDLTNARGSNKSLAANPEILIESIPAKKGIDIN 430
 DB 852 DKLEKSL-KDRESNVLDIQRLK-AEKETLKAMYDSKDALFAOKRELEIRVDEMESEIDE 909
 QY 431 VK-SPEETEKDNTENDSKAEKNREFEED-----QSLKDSDD-----KTPDDDPDQG 479
 DB 910 KKLALENIQNOKRSYEEVVRDLLEELQEKLRNTLEKTKKYEELMEEMKRVNDGSGDT 969
 QY 480 KSEVDFKSE--KSNGLSESS--PGAGKAGSGSTRITR----- 514
 DB 970 ISRLKIKDELQKEVEELTESPSESCKDKGVLEKTRVALQSELDLIVRLDSETKKSEL 1029

QY 515 LRNPDSKLSQKSOYAAAAAHEANKLFKEGKEVLYVNGGEISRLSTK--KEVIKNGIN 572
 DB 1030 LRQKKKLEELKQVOZALAEPTAAKLAQEAANKL--OGETLELKEKNSEVTASNV- 1085
 QY 573 NYFKLOEGKRYRYHNOYSTNSPALNKHQHRREDHDKRRHLAHCFLTTPAGEFKMNG---- 628
 DB 1086 -----EKSKTLESQLVAVANNMELDEBKKNRDALEKKAKALDAMLEEMDQLESTGEGKK 1139
 QY 629 SVHGSKVL-----ITSTRLTIITOLENNIPSSFLHPWASHRANWIRAYOMCSKP 678
 DB 1140 SLYDLKQKESDMELRMOISELOSTIAKLEK-----IKST----- 1175
 QY 679 REFALALALECAVCPVVMPLIWRFLGHTRLHRMTSIREEEKYKKKKOE--DEE 735
 DB 1176 -----LEGV-----ARQGELEAQLAKSNVEKOKKKVEILDLEDK 1211
 QY 736 TMOQA--TWVKYTFPVKHQVMKQKGEYRVYGVGSMWISTKTHYRVFKLPNGTNVN-- 791
 DB 1212 SAQLAEPTAAKQALDKTKKLEQELSEVO-----TOLSEAN-----NKNVNSD 1254
 QY 792 -YRKSLEGKNNMDEMMSDKRCSRKIKIPEDSKEDYKGSDAKAGDOME--M 847
 DB 1255 STNKHLETSPNNKLELEAEQAKQALEKRRGLSEELKHVNEOLEEEKKOKSEKRV 1314
 QY 848 DISK-ITTEKRDYVELLDSDSDKPCKEEPMVEVDDMKTESHVNCQESSQVDPVNVSEGF 906
 DB 1315 DEKEVSELKDQLEEVVASKKAVTEAKN-----KKESELDELKQYADVVSBRKS 1365
 QY 907 --HRTSYTK--KTKSKLDGLER-----RIKQFTLE-----EKORLEKILEGGI 949
 DB 1366 VBOULTQAKNELNRLNTAEAGQDLRAERSKKAEFLDEAVAKNLEBETAKVAKRAM 1425
 QY 950 KGI-----GKTSNKKLSSES-----PVTKRKECCQSSMQBOESP 987
 DB 1426 KAAETDYRTKSELDAKAVSSEQYVQIKRLNEBELSELRSVLEEDERONS-AIAAKTTA 1484
 QY 988 NANNQPEDLDIGCSQSSSVLRMS-----PFSHTTNK--LYPKDRVLD 1030
 DB 1485 ESALLESILKDEIDAANNAKAKAEKSKLELVAYAELESLEKSGVAVNEFLRKNAELDD 1544
 QY 1031 V-----SIRSEPIK--CPKO-----NSTENDIEKVSDLASR 1060
 DB 1545 LNARLDRETESRIKSDKEKKNTKQFADLEAKVEAEQREVVYTIIRLKKLLSDIIDLSTQ 1604
 QY 1061 GOEPTKSKTKGNDFTLDSK-----LA-----SADDIOTLCKKKKPLIOR-- 1101
 DB 1605 LDTETKSRK-----TEKSKKLEOTLAEERRAEEGSSKADE-----ETRKQVOMEVD 1653
 QY 1102 ESDTIYSSSKSALHSS-----VPKSTNDR-----DATPLSRADFE-----G 1138
 DB 1654 ELRAQLDSEBALNALNSEKKIKSLVAEVDVEKQLEDELILADKLVKAKRALEVELEVRD 1713
 QY 1139 KLGCSSESNTLENSD--TYSIODSSEDMIVONSNSISEQFTRQDVEVLEPLKE 1196
 DB 1714 QLEEBEDSRSELDSKRRLLTLEVEDIKKKYDAEYQNTKLEAKKKLLDQVDVLK----- 1768
 QY 1197 LVSGSESTNCEDRLPVKGTENGKKRPSQOKKLEEPVKKCSQDKIKNTTJKKNENRES 1256
 DB 1769 -----KLEDEKKKLLNSESERAKKLEBENEDPL-AKIDAVAKNRSRAEK 1811
 QY 1257 EKKQORTFOINGDKNPKIYLKGECLKEISETESRVVSGNVEPRVNNIKITPNDIKSL 1316
 DB 1812 DRK-----KYEKDLKDYK--LNDEAATK--TQTEIGAAKLEDDIDELRSKLEBOQAKAT 1863
 QY 1317 TVKSALRPFLNGVIMEDENERNSETKSHLSSSDA--EGNYRDSLETLTPSTRQSDTQ 1375
 DB 1864 QADRS--KTLLEGIDNLRQIIEDEGKIKMRLEREKRALBELEBELRETYEAEADSKS-- 1919
 QY 1376 TTTPSASCPSESNVQYEDMEITSE--YKAYSSPITSSEENLSNDPIDENG-LPIN 1431
 DB 1920 -----EAGSKRLVELELEDARRNLQLEIDAKLEIAEDAKSNLQREIVEAGRLLEE 1970
 QY 1432 KNEVNGESKRRKIVITEVTTWTSTVATESKTVIIVEKGDKQTVVSTENCAC-----STV 1486

DB 1971 STARTSDRSRKLAEIALTFAQVDAEQAKNOQIKENKIELEKVRKKFGSESEKTK 2030
 OY 1487 TTTTIVTKLSTPSTGSGVILISVKPOSKTV 1517
 DB 2031 TKEFLVEKLETDYKRAKKEADEQOORLTV 2061

RESULT 11
 ANK3_HUMAN STANDARD; PRT; 4377 AA.
 ID ANK3_HUMAN
 AC Q12955;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Ankryrin 3 (ANK-3) (Ankryrin G).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=95138209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.;
 RT "Ankryrin. A new ankryrin gene with neural-specific isoforms localized
 at the axonal initial segment and node of Ranvier.";
 RL J. Biol. Chem. 270:2352-2359(1995).
 CC - FUNCTION: Membrane-cytoskeleton linker.
 CC - ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 alternative splicing.
 CC - TISSUE SPECIFICITY: Expressed in brain and other tissues.
 CC - SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC - SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13616; AAA64834.1; -
 DR HSSP: P55273; 1B18.
 DR Genew: HGNC:494; ANK3.
 DR MIM: 600465;
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000488; Death.
 DR InterPro: IPR000906; ZUS.
 DR Pfam: PF00023; ank; 24.
 DR Pfam: PF00531; death; 1.
 DR Pfam: PF00791; ZUS; 1.
 DR PRINTS: PR01415; ANKYRIN.
 DR SMART: SM00248; ANK; 21.
 DR SMART: SM00218; ZUS; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 21.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 DR Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
 KW REPEAT 73 102 ANK 1.
 FT REPEAT 106 135 ANK 2.
 FT REPEAT 139 168 ANK 3.
 FT REPEAT 172 201 ANK 4.
 FT REPEAT 203 230 ANK 5.
 FT REPEAT 234 263 ANK 6.
 FT REPEAT 267 296 ANK 7.
 FT REPEAT 300 329 ANK 8.
 FT REPEAT 333 362 ANK 9.
 FT REPEAT 366 395 ANK 10.
 FT REPEAT 399 428 ANK 11.

FT REPEAT 432 461 ANK 12.
 FT REPEAT 465 494 ANK 13.
 FT REPEAT 498 527 ANK 14.
 FT REPEAT 531 560 ANK 15.
 FT REPEAT 564 593 ANK 16.
 FT REPEAT 597 626 ANK 17.
 FT REPEAT 630 659 ANK 18.
 FT REPEAT 663 692 ANK 19.
 FT REPEAT 725 758 ANK 20.
 FT REPEAT 729 758 ANK 21.
 FT REPEAT 762 791 ANK 22.
 FT REPEAT 795 825 ANK 23.
 FT DOMAIN 1519 1898 SER-RICH.
 FT DOMAIN 4090 4174 DEATH.
 SQ SEQUENCE 4377 AA; 480399 MW; F42379E55768B684 CRC64;

Query Match 2.6%; Score 374; DB 1; Length 4377;
 Best Local Similarity 17.2%; Pred. No. 5; 3e-06;
 Matches 526; Conservative 420; Mismatches 1080; Indels 1038; Gaps 116;

OY 17 QDSEDDDEDEMEEDDDSDY-----PEEMEDDDDA----- 47
 DB 1441 KETESDODDEIEKTRQROSPASLAKRKRYSLTEPGKITERGTGATNSLPPTYSTKPPFT 1500
 OY 48 ----SYCT-----ESSFRSHSTYSYTPGRKRP-----RYHRKPSPILEEKDIPPLE 89
 DB 1501 RPYQSWTATATVPGPAKSGFTSLSSSSNTSPASPLKSLWSVSTP-SPIKSTLGASTS 1559
 OY 90 PPKSSEDLMPNEHIMNVIAIYEVLRNEGTVLRSPREFDPCALYSQOQTLMAEMHV 149
 DB 1560 SVKSIDSVASP-----IRSLRTMSSPIKTIYVSGSPYNIQVSSQTILARAVTEAPPLKG 1613
 OY 150 VILKAVIREEDTSNTNPG-----PADLKOSVNSTVTFIDGMTPEVLRYVCES 197
 DB 1614 LASNSTPSSSTPVTTAGSLLEKSSITMPRPASPKNTNM----- 1653
 OY 198 DKEYHNHLPYQ-----EAEADYPYGPVENKIKVLOFLYDOFLTTNIAREELMSEGYQYDH 253
 DB 1654 ----YSSLPKPSITTSAPLIPSSPLKSVSPKSRVDISSAKITMASLSLSPVKQMPGH 1710
 OY 254 CRVCHKIGDL-----LCETCSAVYHLECYKPPLEEVPEDEMOCEVCVAHKPGVT 304
 DB 1711 AEVALVNGSISPLKYASSSTLINGCKATATLQ-----EKISSATNSVSSVSAAT 1760
 OY 305 DCVAEIDGN-----KYIRHEPIGYDRSRK-----YWFILNRLLIEEDPENENEK 350
 DB 1761 DTYEKFVSTTTAMPFSPRLKRYVSAAPFOSLTPPSASALYTLGSSISATTSVYSSII 1820
 OY 351 KIWYYSF-----KVOLAELIDCLDKDYWEAELCKILEEMREETH--RHMDITEDLTNKR 403
 DB 1821 TVPVYSVNVNLPALPKLPLDSDNSFTKSAALASPICKTLTETHNPQHPFRTSSPV---- 1876
 OY 404 GSKKSFLA-----ANDEILIESIRAKKGIDNVKS-----PEETE 438
 DB 1877 -KSLFLAPALKLTSPSSLSSSQELIKDYAEKEDIMRTALITQTDVPEKKPFQDELPK 1935
 OY 439 KDKNEPNDKDAEKNPEEEDDS--LEKD-----SDDK--PPDDDEPGKSE 482
 DB 1936 EGRIDDEPKYIEKVEDLVKVSLEIKKDYCVDNKSPSPSPSDGHSPEDEMIETSSSE 1995
 OY 483 VCDPKSEKSNGLSESG-----AGKGS-----GSTRIITRLRN--DSKLSOLK--- 526
 DB 1996 --EIREARQQAASQSSLSLPERQYAKAKASEKDYNTKYVIDVLTNDIGSSSLNLNLYKF 2053
 OY 527 -----SQOY--AAAHEANKLEFKEGEVLYVNSQGEISRLSTK---KEYIMKGI 571
 DB 2054 EDAAKDGEGQKRVLPALALQEHKLMKMPASMRSTSEKELCKMADSFEGTPTILLES-- 2111
 OY 572 NNYFKLQGEKGRVYHNYQVSTNSFALNKHQREDHOKRRLAKFKCLPAGGEKMGVSYH 631
 DB 2112 PDPFSOHDDQSPSLSDSGFTSEKTPPSAPQSAETTPGKPLFHEVLPVPTITTRIEVYH 2171
 OY 632 GSKVLTISTRLTITOLENNIPSSFLHPNMASHRAWMKAVQMCSPREPALALILECA 691


```

Db 2172 VIRSDPSAGVPOTOPPEEPV-----SPKP-----SPTFMELE 2204
QY 692 VKPVVMLPIMREFLGHTBLHMTSTIEREKEKVKKKKKOEDEEMOATVWKTTPFKH 751
Db 2205 PKPTT-----SSI-----KEKVAFOKKASSSEED-----DH 2230
QY 752 OVWKOQ-----EEXRVYGYGWSMISKTHYVRYVPKLPQNTNVYRKSLGCTKNNDENM 807
Db 2231 NRVLISKGMRVKEETHIT-----TTTMYVHSP--PCG-----EASERIEETM 2271
QY 808 DESKRCOSRSPK-----KIKIEPD-----SEKO-----E 832
Db 2272 SVHDMKAFQSGRDPSELACLPFEKSAVSPDVHKSAETSAOHAEDNOMKPKLERITE 2331
QY 833 V--KGSDA-----AKGADONEMDI--SKITEKKODVKEHLLDSODK--PCKEPEME 878
Db 2332 VHIKGNQOAEPTVETIHTKHKHPEKEMVYVQKDSLRODINKDLPRKHDAFPCSEDOQ 2391
QY 879 VDDDKMT--ESHVNCQSSOVYVNVBSGPHLRTSYK-----KYKSKT-----921
Db 2392 QEEBELTAEBELPBYLBSRYN--TPVSOEDSRPSSAQLISDDSYKTLKLQHSIEYHD 2450
QY 922 DGLERRIKOFTLEKORL--EKIKL-----EGSIGIKGTSTSSKNLS 964
Db 2451 DELSELGCESTRFAEKMLSEKLDVSHSDTEESTDHAGPPSSSELQSDSKRSREKIATAP 2510
QY 965 ESVYITAKEGCQSDSKROEOPNANDOBEDLIQCCSOSDSSVLRMSDPSHT-----1017
Db 2511 KKEILSKIKYKVSNGVGK-----VSKDEHDKVTVLHYSGN-----VSSPKHAMMMHFTFE 2561
QY 1018 -----TNKLYPKRVLDVDSIRSPETCKCKONSIENTIEKYSDLASROEP--TKSKTK 1070
Db 2562 DRLDRGKEKLIYEDRV--DRTVKEAEKLETVESOFFRDKTEKLN--ELQSPKEKKARPK 2616
QY 1071 GNDFFIDSKLASAD--IGTLICKKK-----PLIOEEDTLVSSKSA 1113
Db 2617 NGKRTYSQOSTSPSEKVLITELLASDENVKAROHGPOGQGFPAKAEKAPSLPSSPEKM 2676
QY 1114 LHSSVPEKSTNDROT-----PLSRAM-----FEGKLGCDSESN 1147
Db 2677 VLS---QOTEDSKSTVAKGASISQSKAPDPOGFOIKQSKLSIRLKPEOGHNAKSKDM 2733
QY 1148 STLSSNDJVS-----IQOS-----SEEDMAYONSNSISBQ-----FRPREOD 1186
Db 2734 SQEDRSKDGOSRILPVKKIOLSKLPYQVFAREKQAKAIDLPRDESYSVOXDFMVLKTKDEN 2793
QY 1187 VEYLE-----PLKCELVGSESTGNCEDRLPYK 1213
Db 2794 AQSENEIYVNDGSDNYKKORTEMSKAMPDSFEQOAKDLACHITDILATRGWMDKVR 2853
QY 1214 GTANG--KRPQOQKLEERPVKCSDQIKLAKTTDKKNE--NRESEKKGORTSTFOIN 1269
Db 2854 TWESSGATNNKSOKEKLSHLVH-----DVRENHIGHPRESKVDOKNEFMSVAT 2901
QY 1270 GKDNKRIYKJGCLKEISESRVSGNVEPKVNNIKIIPENDIKSLTYKESAIRFING 1329
Db 2902 ERRERK--LITNLSIEIKEMYKSPS-----KKVLYREVYKGGDHP 2941
QY 1330 DVIMEDFERNNSSETKSH-----LSSSDAGNRYRDSLETLPTKESDSTQTTTP 1379
Db 2942 GGLDOPSRSESSAVASHIPVRADERRMLSSNIIPDGFCEOS--APFKHELQSKLSQSSM 2999
QY 1380 SASCPSPSNVNOVEDMEIETSEKKYT-----SPTITSEB-----ES 1416
Db 3000 SKETVTOHFNSIETDEKVTYSLSKYSKQSYVGLCPLELETSTPTKSPDSLSESPGKE 3059
QY 1417 NLSNDEIDENGLPINKENW-----NGESKRRTVITTEVTMTSYVATESKVIIVYKED 1470
Db 3060 SPSSDVFDHS--PIDGLELAPLAQTEGKEKTLTPVY--SFVQVQKQYKEKELOQOG 3113
QY 1471 KQTVVSTENCARSTVTTTTVTKLSTPSTGSDYDIISVKEQSKTVVTTVTDTSLTTTG 1530

```

```

Db 3114 VKRILISOECTVQETRGNFYTTTRQKQRPSPQSGEDDTLEQVS-----FLDSSG 3163
QY 1531 GTLVTSMT-----VSKYSTRODKYKLMKF-----SRPKTRSGALDSYK 1571
Db 3164 KSPILPETPSESEVSYETSTPDSLAVIPGRSPPIPEVSESESEEOAKSTSL-----3218
QY 1572 FVTSTKRSIFV--LPRNDLKLARKGIREVYFYNKRPALD-----1613
Db 3219 --KOTVIEETAVENEMENDVSKDSNORPKNNRAVIEFPPLPLDQDIESDKKHHNYLPE 3276
QY 1614 -----IMVPSP-----RPTGRTWRYR 1631
Db 3277 KEVDMIEVNLQDEHDKYQALAEVYIRVOP--PSVPBGADVSDDSDDESITQVPVVKKYTRK 3335
QY 1632 LQTV-----KSLAGVSLMLKLYMAS-----LRDDMAAVYPPGSGSTRITETSE 1674
Db 3336 LKEYDDQOKRPKYASAKASNOKELESNGSKDNBERGLGIDSPONEIATONGNNDOSITRC 3395
QY 1675 TEITTEILRRD-----VGPYGIREF-----YCIRKILICPT 1706
Db 3396 SIATTAFFSHDDATELSDGLDGLDDEDDGLSDSKLPIQAMEIKDKIOMNTEGLIKRA 3455
QY 1707 G-----VPETKETPTPOKGLRSALRPKRPEPTKQOTGVYIETWAEELIEM 1756
Db 3456 DRFSFSKLEVEIEEGKVGDDEDKPPSSSSSEKTPDKQKGAQFTLLEGNHPRSYF 3515
QY 1757 EIRAFERVEKEKAQAVEQA-----KKRLQOKPTYIATSTS 1795
Db 3516 PDIYFSKYVDEEATPPTKYATATGLDPDWSNNRGDEVDKSRDETKPGLAAEDRS 3575
QY 1796 PITSSTSTISPAQVNAVAPISGVTGTQKVLTKVGSFATVTFQONKPNHOTFATVWQ 1855
Db 3576 P--ATPDTTPAR-----TPDESTPTISENPPFHHGKMEHT-----3612
QY 1856 GQSNQGVNOQKVLGILISSTGSOQTTSQOTTSFOFPAITYITRPN-----SGSGGTTSSNQ 1911
Db 3613 --RSGAIDM-----SKRDEY--EERLQFOIGEBHTSEKSGOGEQSDKSM 3653
QY 1912 VITGPIRPGMTVIRPPILOSTLGKALIRPVNVGCGAPQVWTOIIRQOPSTAVSAPN 1971
Db 3654 VYATRPQOSGDTTVEITNERN-----VETPT--VEP-----NPSIPTSGEOE 3694
QY 1972 TVSTPGQKSLTATSTNSIQSSASQPPRPOGOVYKLTMAQLTOLLOGHGNOGLTVIY 2031
Db 3695 GTSSSGSLKESAAATNTSKYDEPLRTP-----IKMGISAST-----MTMKKE 3736
QY 2032 GOGOTGQOLILPOGVTVLRGCGOOLMAAMNGVQRLFLPLATATATATTTTST 2091
Db 3737 GREITDKTEAV-----MPSCOGLENETITMISN 3765
QY 2092 TAAGTEGQOSKLSIPQMVHODKTLRPAQSSSVGPAKAOPOTAOPSARPOP--TOPQSP 2149
Db 3766 TA-----NSQAGVPRHEHNDOKONE 3786
QY 2150 AQPEVOTOPREVOTQTTVSSHVPSERAPTHAOSKPOVAQAQSPQSNVQSGFVRVQSPQ 2209
Db 3787 NNNNNLDSSTIOTDINMSIVLTE-----HSAPCTCTEKDNPVKVSSGK 3831
QY 2210 TRI-----RPSRPSQLSPQOSQOVOTTSQRIPIOP--HTSLQIPSGOP 2252
Db 3832 TGVLOGHCVRKQKVLGEOQKTKELIGIKOSKLDLKAATSPDITPRPHMSTYKASKKO 3891
QY 2253 QSOPOVQSTOTLSSGQTLNVOYVSSPSRPOLOIOPOPOVAVAPLOQOOVOLSQISQ 2312
Db 3892 VQOSKRTKALTITSCVDVKSRIPVKN-----TPRDNIIAV-----RKA 3929
QY 2313 VVAQIOAQ--QSVPOQIKOLPIQIOQSSAVOTHOIQWNVTVQASVQ-----2359
Db 3930 CATOKQGPENKGAOLPSKLPVKVRSCTVTTTTTATTTTTTTTTTTTCTVVRKSQ 3989
QY 2360 -----EOLORYOOLRDQOQKQKQOQIIEKREHNLQASNOSEIIOKOVVM 2403
Db 3990 KEVCHSIEIFYKGISGETKLIVDRL--SEERKMOSELSDEESTSRNTSISE-----4040

```



```

QY 2404 KINAVIEHLKOKKSMTPAREENORMIYCONVMKYILDKIDKEKOAKRRKRESEVOK 2463
DB 4041 -----TSGGQGPSTYTKSAR-----DKTEAPLPLKSEKAGSER 4075
QY 2464 RSKQ 2467
DB 4076 RSR 4079

RESULT 12
MLP1_YEAST
ID MLP1_YEAST STANDARD: PRT: 1875 AA.
AC 002455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR05W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Genet. 237:359-369(1993).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L01992; AAA34783.1;
DR EMBL: X73541; CAAS1948.1;
DR EMBL: Z28320; CA82174.1;
DR PIR: S38173; S38173.
DR SGD: S0001803; MLP1.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match 2.5%; Score 360; DB 1; Length 1875;
Best Local Similarity 17.9%; Pred. No. 6.5e-06;
Matches 316; Conservative 298; Mismatches 656; Indels 494; Gaps 72;
QY 7 BEEDGAEFTQSEDEDEDEDDSDYPEMEDDDDDASCTSSFRSHSTYSSTPG 66
DB 314 KEELNSIRLNTAKYADDSKOTPENEDLKLKELTKKLAQCEKCR 363
QY 67 RKKPVHRPSPILEKDIPLLEF-PKSESDLMV-----PNEHIMNIAIYEVLR 115

```

```

DB 364 -----LSSITDDEADENENI,SAKSSSDFFELKKOLIKERRTKEHONQIETPIV-- 412
QY 116 NGETVLRLSPFREDPCALYISOECTLMAEMHVLLKVLREEDSNTPFGADLKDSY 175
DB 413 -----ELEHKVPITNSFKERTDLENE--LNNALLLEHTSNEKNKAVKELNKK 459
QY 176 NSTLYFDGMTWPEVLRYVCESD-----KEYHNVLPYQEAEDYPYGPV-ENKI 222
DB 460 NQKL-----VGCENDLQTLTKQRDLCKHQIYLLITLNSVSDSKGFLKKEI 506
QY 223 KYVLQFLV--DQFLTNIAEELMSEGVIOYDHCRVCHLGLL----- 264
DB 507 OFIIONMOEDDSTITESDSQKVTERLVFKNTIIOLEKNNAELKVVRLADLKSEK 566
QY 265 -----CCENCSAVYHLECVKPPLE-----EVP-EDEMQCEVY 296
DB 567 SKOSLOKISSEYVNEAKEAITLTKSEKMDLESTIELOKLEBELKTSVPNEASYSNVI 626
QY 297 AHKVPQVTCVAEIQ-----KNKPYIRHE-----PIGYDRSR 329
DB 627 KQLETKRDLQESQVQDLQIRISQITRESTENMSLKEIQLDYDSKSDISIKLGKESSR 686
QY 330 -----KYFPLNRRLLIEEDTENENKKIMY-----STKVQDLAELIDCLK-DYW 373
DB 687 ILAEERFKLLSNLIDLTK-AENDQLRRFDYLONTILKODSKTHETLNEYVSCSKSLIV 745
QY 374 EMLCKILEEMREIHRHDITEDLNNKAGSKSP-----LAANEELISIR-- 422
DB 746 ETELLNKEEQKLRVHLEKNKQEL-NKLSPEKDSLRINVTQLOTLOKEREDELLETRS 804
QY 423 -AKGCD-----IDNVKSP-----EETEKDKNE-----TENDS 448
DB 805 CQKIDELDALSELKETSQKNDHNTKOLEEDNNSNIEMWQNKTEALKKDYESTIVSDS 864
QY 449 KDAEKREEFEDOSLEKDS-----DKTPDDP-----EQGK-----SEVG 484
DB 865 KQPDIEKLQYKVSLEKEIEEDKIRLHTYVMDETINDSLREKLEKSKINLTDAYSQIK 924
QY 485 DFKS-----EKSNGELSESPGAGKASGSTRITLRLRN-DSKLKQSQVAA 534
DB 925 EYKOLYETTQSILQOTNSKIDES--FKQFTNOIKMLDEKTSLEKISILKQ----- 975
QY 535 HEANKLEKGEKEVLVNVNDSOGEISRLSTKKEVIMKGNIN--NYFKLGQEGKYRVYHNOYS 591
DB 976 -----MNLNNELEDLQKKGKEKADFKKRISILQNNKNEVEAVKSEYSKSLKINDD 1030
QY 592 TNSFALNKHQHRDHDKRLHAKFCLTPAGEFKWNGSVHGSVLTITSLRLTITTOLENN 651
DB 1031 QQTIVANTQNNNEOELQKHADVSKTISELRE-----QLHTYRG-QVKTLLNSRDQLENA 1084
QY 652 IPSFPLHPMASHRAMIKAVOMCSKPREFALA-----LAILCAVAPV----- 695
DB 1085 LKEN--EKSNSQKESLELELDSNRIEDLSQNKLLIDQIITYAADKEVANNSTNGRG 1142
QY 696 -----VMLPIWREFLGHTRLHMTSIEREEK--EKVKKKEKQEBEETMOQATWVYTF 747
DB 1143 LNNILTLRREDIILD--TKVTVAERDAKMLRQKISLMDVDELQDARFKLDSRVE-- 1195
QY 748 PYKHQYWKQGEERYVTGIGGWSMISKTHYRFPVPLPGTNNVNYKSLGRTNANNDENN 807
DB 1196 KENHSSIIIOHDHDI-----MEKLNQNLNR-----ESNTLNLNELENNNKKREIQ 1241
QY 808 DESDKRCKSRPKIKIIPDSEKDEYKGSAAAGADONEMDISKIT-----EKKDOVKEL 863
DB 1242 SELDKLQNVAP-----ISELTLALKYSQEK--EQELKLKLEEVHNRKRSQDILE- 1291
QY 864 LPSDSKPCKEEPMVEVDDMKTESHVNCQESSQVDVNVSEGFHLTYSYKKTYSKSLDG 923
DB 1292 -----KHQDLSSSDYKLESEL--ENLKEELENE-----RQGAABEKFRMLRR 1334
QY 924 LERRIKQFTLEEKQRLKLEGGIKGIGKITSNNSKULSESPIVTKAKKQCSQSMRO 983

```


Dh 1335 QAOERLTKSLSDSLTEOV-----NSLRDAKNVLENSL-----SEANARI 1375
QY 984 EGSFNANNDPDPDLIGCSOSDSSVLRMSDPSSHNTNKLYPEKRDVLDDVIRSEPTKCPKQ 1043
Dh 1376 FELONAK-----VAQGNQNLLEAIRKLOEDAERKASRELQAK-----LEESTSY 1418
QY 1044 NSTENDIEEKVSDLASGQPTSKTKGNDFFIDDSKLASADTGLICKNKKPLQOES 1103
Dh 1419 ESTINGNEIEITTL-----KEEIEKQROIOOOLQATANE-----QNDL 1457
QY 1104 DTIVSSSKSLHSSVPKSTNDRATPLSRAMDEGKLGCSSESSTLENSSDTVIODSS 1163
Dh 1458 SNIVESKAKKFEEDKIKFTEKTOEVNEKILQEOERL--NOPSINNEIKKKWESHEQ 1515
QY 1164 EEDMIVONSNEISIEOPR--TRQDVEVLPLKCELVGSGSTGNCEDRLPVKGTANG-- 1219
Dh 1516 EVSQIKREAEALKKRIPLTEEKINKIIEKKKEEL-ERKEFEKVEER--IKMEQSGEI 1572
QY 1220 -----KPSQOKKLEERPVNKCDOI-----KLNKTTDKKNNE 1254
Dh 1573 DVALRKQLEAKVOEKQELNENYKRIQOEELKDPVHSSHISDDEBKLRALIEISRLKEEF 1632
QY 1255 ESEKKGORTSTFOINGKDNKPKIYLGECLK-----EISRSRVSSGVEEKVNNINK 1306
Dh 1633 NNELOAIKKKSF-----EGKQOAMMKTTLLERKLAKMESOLSETKQSAESPXSNNV-- 1686
QY 1307 IIPENDIKSLTVK-----ESAIRPFGVDVIME-----DPENRSSEKSHLSSSD 1353
Dh 1687 ---QNPLLGLPKLEIENSNSPFNPLLSGEKLLKLNKSSSGCFNPFTSPKNNHLONDND 1743
QY 1354 AEGNYRDSLETLPSTKESDSTOTTPSASCPESNS--VNOVEDMEITSEVKVYTSPP--- 1409
Dh 1744 K-----RESL-----ANKTPPTLLEPSFNIPASRGLISSSTLSTPNDIELTSNNPAQ 1794
QY 1410 -----ITSEESNLSDNFIDENGFLPINKENYNGESK-RKYTYTPTW-----TSY 1456
Dh 1795 DSSNRNVQSGEDTEKKE-----GEVYKKEALIEEOTKSKNRPIDVEGELKDNEDDTTEN 1849
QY 1457 ATESKTVIKYKGGKQOTVVSSTEN 1480
Dh 1850 INESKKIKTEDEBEKETDKVNDEN 1873

RESULT 13
ATTRX.HUMAN
ID ATTRX.HUMAN STANDARD: PRT: 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATTRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATTRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
RP VARIANTS ATTR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATTRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATTR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lissi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Coliaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATTRX gene encoding
RT a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Strayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Gonzalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Gonzalez G., Villard L., Strayton C.L.,
RA Millaudeau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATTR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (Attr-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RL Submitted (Dec-2000) to the EMBL/Genbank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timst S., Villard L., Khrestchatskiy M., Fontes M.,
RA Coliaux L.;
RT "Specific interaction between the XNP/ATTR-X gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATTRX) at
RT pericentromeric heterochromatin and the short arms of acrocentric
RT chromosomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
RN [9]
RP DISEASE.
RX MEDLINE=20213147; PubMed=10751095;
RA Villard L., Fontes M., Ades L.C., Gecz J.;
RT "Identification of a mutation in the XNP/ATTR-X gene in a family
RT reported as Smith-Fineman-Hyers syndrome.";
RL Am. J. Med. Genet. 91:83-85(2000).
RN [10]
RP VARIANT ATTR-X SER-1713.
RX MEDLINE=97196774; PubMed=9043863;
RA Villard L., Iacombe D., Fontes M.;
RT "A point mutation in the XNP gene, associated with an ATTR-X phenotype
RT without alpha-thalassemia.";
RL Eur. J. Hum. Genet. 4:316-320(1996).
RN [11]
RP VARIANT JM GLN-2131.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Gecz J., Matel J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [12]
RP VARIANTS ATTR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aflimos S., Asenbauer B.,

Db 527 DIFENLETAMEVOSSVDHQDSSS-GTEOEY-----ESSS 560
QY 595 FALKHOHREDHRRRLAHKFCUJTPAGEFPMNGSVHSGVLTISTLRTITOLENN--- 651
Db 561 VKLN-----ISSKDNRGKIKSKTT-----AKYTKLYKLPVSLPNSPIK 601
QY 652 -----IP-----SSFLHPNMASHRAMIKAVOMCSKPREFA-----LALATLE 689
Db 602 GADCEVPODKGYKSCGLNPK-----LEKCGLOEJNSDNEHLVENEVSLLEE 650
QY 690 CAVK---PYVWLPJWREFLHTRLHMTSTIERE---KEK-----YAKKEK----- 729
Db 651 SDLRSPRVATTPLRP-----TETNPVTSNSEDCEKNETVREKOKLSVAVPKKCKRNSSDS 706
QY 730 -----KOEERTMOQ-----ATWVXYTFP 748
Db 707 AIDNPKNKLPKSKQSTVQNSDSDMLAILKGVSMHSSSSDDIDINEIHNHKLTYD 766
QY 749 VKHQVWK-QKGEERYVGYGWSW-----ISKTHYRFPKLPGMTNMY----- 792
Db 767 LKTQAGKDDGKRRKRSSTGSDFTKKGSAKSIISKRRQTOQSESSVYDSELEKEIK 826
QY 793 -----RKSLGCTKNMDENDSDKRR--CSRSPKIKIIEPDSKDEKVGSDAA 839
Db 827 SMSKIGARTTKKRIPTK--DFDSSDEKHSGKMDNQHKRLKTSQEGSSD-----DAE 880
QY 840 KGADQNMDSITTEKKQDVKELLDSDSPCKEERPEVDDDMKTESHVNCQESSVDVY 899
Db 881 RKQERETFSAGEVQDQDTIMELRDR--LPKKQOASATDQVDLKSQE-QSFTSLEY 936
QY 900 VNNS-----GFHRTSKTKTKTSKSIDGLE-----RRIKQPTLEE 936
Db 937 RKVAETREKSKHLTKTKCKVQ---DGLSDIAEKFLKKQSDSETSDDKQSKGTEEK 992
QY 937 KQ-----RLEKIKLE-----GGIKGIGKSTNSKLLSSPIV 969
Db 993 KKSQDPKKVYKMQEYESSDQTEKLPEREICHPKGIKQKNGTQGEKKSKIRDK 1052
QY 970 TKKECCQSDSMKQESSPANNDQPED-----LQ-----OGCSQSDS 1006
Db 1053 TSQKDELSTYAKSTGKSGSDSSDEKSKNGAYGREKKRCKLCKSSRKRRQDSSDT 1112
QY 1007 SVLRMSDPHTTKILPKRVLDDVSIKRPETKCPQNSIE-----NDIEKYSDLAS 1059
Db 1113 EKXSMKEDGNS--DKRLKRIELRRRLSSKRNKTEQSGSSSDAEESSDNKK 1167
QY 1060 RGQEPKSKTKGNDPFIDSKLASADDITLICKNNKPIQESPTIVSSSSALHSSVP 1119
Db 1168 KKQR-TSSKK-----AVLYKEK-----KNSLRTSTK 1194
QY 1120 KSTNBRDAPFLBRAMDEKGLGCDSESNSTLNSSDTVSIQD-----SSEEDMIYONS 1172
Db 1195 RFOADITS---SSSDIE-----DDQNSIGEGSSDEQKIKPVTENLYVLSHTGQSSG 1246
QY 1173 NESISEOFRTRODVE-----VLEPLKCELVSGSTGNCERLVLVKGTEANG 1219
Db 1247 DEALISGVVYVDDDDDDNDPENRIAKKMLLEIKANLSSDD-SSDDE-PEEGKKRTG 1304
QY 1220 KNPQOKLLE-PPVNNKCSO-----IKLKNY-----DKNNENRES 1256
Db 1305 KONEENPGDEEAKNOVNSDSDESSEKPRYRHLRLKRLVLSGEGSEKKTPEKREK 1364
QY 1257 EKKQRTSTFOJNGKDNKIKYLKBECLKEISERASVSGNVEPKYNNINKIIPENDIKSL 1316
Db 1365 EYKGRRR--KYSSBDSBDSDFQESGVSEVSES--EDEQRPFRRAKKALEENQSY 1419
QY 1317 TVKESAIRPFINGDIYMEDFNERNSEFTKSHLSSSDAAGNRDLEFLPSTKES----- 1371
Db 1420 KKKKRRRLKQVEDSSS-NSNSSEEEEEKKEEKEEEDENDSDSKSPGGR 1477
QY 1372 -----DSTQTTTPSASCESNSVQVEDMEIETSEVKV-----TSSITS- 1412
Db 1478 KIRKILMDOKLRTETQNAKKEEERRRRRIARERERERKIKRIVITEIDASPKCPIITKL 1537

QY 1413 -----EERSNLSNDFIDENGILPIKNENVNG-----ESKRT-----VITEVTT 1451
Db 1538 VLDEETKEPVLQVQHRMNIKLRPHQVDQPMWDCCESVKTKKSPGSCILAHCMG 1597
QY 1452 MSTVATES--KTVIKVEKGDQTV 1475
Db 1598 LGKTIQVVSFLHTVILCLDKIDFSTAL 1623
RESULT 14
ID TRX_DROVI STANDARD; PRT: 3828 AA.
AC 024742;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z50038; CAA90349.1; .
DR HSSP: P20393; IABY.
DR FlyBase: FBgn0014844; Dvir\trx.
DR InterPro: IPR003889; FYRIC.C.
DR InterPro: IPR003888; FYRIC.N.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001628; Znf_C4steroid.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00628; PHD. 3.
DR Pfam: PF00856; SET. 1.
DR SMART: SM00542; FYRC. 1.
DR SMART: SM00541; FYRN. 1.
DR SMART: SM00249; PHD. 4.
DR SMART: SM00508; PostSET. 1.
DR SMART: SM00184; RING. 2.
DR SMART: SM00317; SET. 1.
DR SMART: SM00389; Znf_C4. 1.
DR PROSITE: PS50280; SET. 1.
DR PROSITE: PS01359; ZF_PHD. 1; 3.
DR PROSITE: PS50016; ZF_PHD. 2; 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator.
FT ZN_FING 1251 1334 PHD-TYPE 1.
FT ZN_FING 1335 1380 PHD-TYPE 2.

FT ZN.FING 1408 1469 PHD-TYPE 3.
FT ZN.FING 1708 1767 PHD-TYPE 4 (ATYPICAL).
FT ZN.FING 1768 1818 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-GLN.
FT DOMAIN 243 251 POLY-ALA.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 538 546 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
SO SEQUENCE 3828 AA; 413721 MW; 3205CF303A3C504 CRC64;

Query Match 2.5%; Score 358.5; DB 1; Length 3828;
Best Local Similarity 19.0%; Pred. No. 1.6e-05;
Matches 683; Conservative 439; Mismatches 1267; Indels 1215; Gaps 178;

QY 5 EEEEDGDAAEFQDDEDEDEDEDDDDSDYDEEM-----EDDDDDASYCT 51
DB 523 EDNDDGSA-----SSDAIEDDEDIDDDAENEENEAASEKSAETTASVDEKEADRDQVW 577
QY 52 ESSF--RSHSTYSSTPGRRKPRVHRPRSPILKEKDIPLEFPKSEEDLWPNPHI-----M 105
DB 578 DNHFLVLPKSTSS-----RIKPNKRLLEVGIGCSKRSRPSDANGKRRKPKYFGLATL 630
QY 106 NVAIAIEVLNFGTVL--RLSPRFEDFCAALVSQEOCTLMAEMHVLKAVLREEDTSN 163
DB 631 PAKCTPRRRRSATATASQKIGKETEFASFATKYNSS-----FVLROPRLQFQDKS 681
QY 164 TTEGPAKLDSVNSTLYFDGKMTREVLAVYCCSDKEHHVLP-----YQEAIDYRYGPV 218
DB 682 RSTVSA--KPLTPTT-----VLPASSAITSAVNLSEFAL 715
QY 219 ENKIKVLOFLVQFLTTNIAREELMSEGYIYQDHCRCV-----HKLGLLCC 266
DB 716 NN-----ANSAVMAASTCAVCSAPVNNKNDAPLARKYTG-VIAC 751
QY 267 ETC-----SAVYHLECVKPRLEVP-----DEMOEVCVAHKGVTDCVAEIOKNRP 315
DB 752 EVCRRKNSHMTKISKLSTPMHNSPSTSAOSGQOLKC-----TD----- 790
QY 316 YIRHEPIGDRSRKRWFLNRLIIEEDTENEEKIKWYSTVQVLAELIDCLDKDYWEA 375
DB 791 -----GGNCS-----ILSLKSQLNKFKL--YKERCKACMLKCLATLQLP 830
QY 376 -----ELCKIL-EMREIHRHMDITDL-----TNKAR 403
DB 831 GHRSLATLIPASMRREVAPKDDCKPELLSPASLFTAPTASSAGTTIKKSSAETAY 890
QY 404 GSNKSEFLAANE-----EILESIRAKGIDINYSPEETE----- 438
DB 891 NSIKSNPLAENNVTFGCTPLLRPAILEKPLFKIGSDNKKAKESKALGSPVSTSEAA 950
QY 439 -KDKNTEMSDAEKNREFEODSLEKSDDKTPDDDEPGSGSEVGDKSEKSNELSE 497
DB 951 VAPGKTTTAKQDKEKARLEAEKPLSPNAKKTTEANTPETODEOPASTTTTVSASSS 1010
QY 498 SPGAGKAGSGSTR-I-TTIRLN-----PDSKLSQ--LKSOQVAAAHEAN-----KLK 542
DB 1011 TSHTSAATNNSOLETTEAANSAVDNLKQRIDLKGRVHKVCSASIVLGQPLATG 1070
QY 543 EEKEVLVNSQ--EISLSTRKEVIMKGINNVYKLGQEGRYVYHNO-----YSTNS 594
DB 1071 DEEBELAAEAPAPATTTTTSPEYIIKRP-----KSPQMOMITDENDNCASCLTPT 1125
QY 595 PALNKHQR-----EDHKRRHLAHKFCITPAGEFWNSV--HOSKVLITLTLTIT 646
DB 1126 EATAEQPAVKSVLESRSSKSNTOGTEAKRTPATSGSSKGKVTTRNATVTVSASSLAT 1185

QY 647 QLENNIP-SSEFLHPMA-----SHRA-----NMIKAV-----QMSKPREPAL- 683
DB 1186 KQRIENIEVSSISSQAAATQSRRLAKEVNRKLALISIDFMENTYDPAEVCQ--TGFGLI 1243
QY 684 -----ALAL-----ECAYKPVYMLPIWREF-LGHTRLH--RMTS-IERE 719
DB 1244 VETVAORALCFGLSGSTGLDPLIFCACCEPBYHQYCVLDEYNLKSHSFEDTLTSLLETS 1303
QY 720 EKEKVKKKKEQEEEFMOATW-----VKYFEPK--HOVVKOGEE-YRVTGYGWS 770
DB 1304 NNACALSAATNTALNLTQRLNMLCPRCTYCTYCNMSSGSKVACQKQKRYHSTCLGTGR 1363
QY 771 W-----SKTHVYRFPKLP----- 785
DB 1364 RLIGADRPILCVNCLCKSCATTKVSKFVGNLPMCTACFKLRKKGNPCIQCKYDNDNF 1423
QY 786 -----GNTNVYRKSLBCTKN-----NMENMDESCK--RKSRSRPKIKIEPDEKDE 832
DB 1424 DLKMECGDCNOMVHKSCEGLSDEQYNLSTLPESTIEFICKKCARCDVSRNKADRMROA 1483
QY 833 V-----KSDAKGADONEMDISKITEK 855
DB 1484 VMEFFKSLYSYLKLSKROACALLKSPKRWKCCSAGQAKAHSQ-----GLQPK 1538
QY 856 KDQVKEILDSDSDKPCKEPMEDD-----DMKTESHVNCQESSQYDVVNVSEGFHLRTS 911
DB 1539 ALQFTYNGIGDSQNSDDIYEFKQHSNRKPSYRVC-----SCLQPLSQSPFSL-YD 1594
QY 912 YKAKTKSKLDGLERRIKOFTLEEKQREKIKLEGGIKIGKSTNSSKNLSESPYTK 971
DB 1595 IKOKIASMAVSLAE-----FNYSQVIOO-----SNCEDELIAVK 1631
QY 972 AKEGCSDMSROBOSPNANDQ-----PEDLIOGCSQSDSVLAMSQDSHNTNLYK 1024
DB 1632 -----ELSSQFPFPOETACTDALEEDMFESCGYEELKESPTTYAEHHTASQAPR 1683
QY 1025 -----DRVLQVSIRESPTCKPKONSIENDIEEVSVDLASGQEPTRSKTK-----GNDF 1075
DB 1684 TGLLDLPIDVDLDG--GC-----AVKTRLDTRVCLFCRRSGSGSLGEBEARLLCGHDCM 1736
QY 1076 IDDSKLASADDIGTLCKNKKPLIOESDPTIVSSSKSALHSSVPKS-----TNDRADT 1128
DB 1737 VH-----INCAMSAVEFEEID--GSLQNVASAVARGMIKCTYCGNRGAT 1780
QY 1129 -----PLSRAD-----FEGKIGCDSESNSTLE--NSDPTVSIQDSEEDMIV 1169
DB 1781 VGCNVKSGEHHYPCARTIDCAFETDKSMYCPAHARNALKANGSPSYTESFV----- 1836
QY 1170 QNSNESISEQFTRQEDVELEPLKCEL-----VSGESTGNCEDR-----LPVKGTEA 1217
DB 1837 --SRPYVELEKRRK--KLIVAPAKVOFHIGSAVAVQOLGIVRFSDSPALYPINFLCS 1891
QY 1218 -----NGKPSQOKLEBRV-----NKCSQIKL--KNTDKNNENRESEKKG----- 1260
DB 1892 RLWSSKEP--WKIVEYVTRTTIQNSYSSTLILDGRNFTVHTPNCSLVOJLAQOLA 1948
QY 1261 -----QRTFOINCKDKPKIYILGECLEKISESRVYSGNVEPRVNNINKIIEPENDK 1314
DB 1949 RMHSLARSOLDLDTDAE--FPNSYVPAD--ENTEE-----EPO-QVADLLPPE--TK 1994
QY 1315 SLTVKESAIRPTING-----DVIM-EDFENRN-----SSETKSHLSSSDAEGYRSLLET 1364
DB 1995 D-AIFEDLPHELLIDGISMIDIFMYEDLGKTELFAMSBQSKOCTYATSAAG--ASVILC 2051
QY 1365 LPSTKESDS--TQTTTPSASCPSNSVNOV-----EDMEIETSEV-KVYTSPP----- 1409
DB 2052 DEDTRNSNSLNKHLVSNCTASNPVDAMLCAARSSQKECEGDVLLKKTDTAPATSPMK 2111
QY 1410 -----ITSEESLNDLFDENGLPIKNKNENVGBEKKRTVITEVYTNMSTYATESEKYI 1464
DB 2112 LDGGSVAARRRRLSKNIAE--GVLLSLNO--RSKK-----EMATVA 2149

RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
 Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 RN [3]
 RP SEQUENCE OF 974-2440 FROM N.A.
 RX MEDLINE=99375328; PubMed=10444336;
 RA Nagaya T., Chen K.-S., Fujieda S., Richer J.K.,
 Horvitz K.B., Lupski J.R., Seeo H.;
 RT "Localization of the human nuclear receptor co-repressor (hN-COR) gene
 between the CMT1A and the SMS critical regions of chromosome
 17p11.2.";
 RL Genomics 59:339-341(1999).
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION. THUS
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
 CC ABSENCE OF LIGAND.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 CC DOMAINS (ID1 AND ID2).
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 CC SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES
 CC FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 CC SPECIFICITY.
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CORN BOXES.
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF044209; AAC33550.1; -;
 DR EMBL: AB028970; BAAB2999.1; -;
 DR EMBL: AB019524; BAA75814.1; -;
 DR TRANSFAC: T04687; -;
 DR GeneW: HGNC:7672; NCOR1.
 DR MIM: 600849; -;
 DR InterPro: IPR001005; Myb_DNA-binding.
 DR Pfam: PF00249; Myb_DNA-binding; 2.
 DR SMART: SMO0395; SANT; 2.
 DR PROSITE: PS50090; MYB_3; 1.
 KM Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KM Coiled coil.
 FT DOMAIN 174 216 COILED COIL (POTENTIAL).
 FT 254 312 INTERACTION WITH SIN3A/B.
 FT DOMAIN 299 328 COILED COIL (POTENTIAL).
 FT DNA_BIND 437 482 SANT-A (POTENTIAL).
 FT 620 670 MYB.
 FT DOMAIN 501 557 COILED COIL (POTENTIAL).
 FT 607 617 PRO-RICH.
 FT DOMAIN 988 1816 INTERACTION WITH ETO.
 FT 2055 2059 CORN BOX OF ID1.
 FT DOMAIN 2263 2267 CORN BOX OF ID2.
 FT 58 64 POLY-GLN.
 FT DOMAIN 593 603 POLY-ALA.
 FT 1032 1035 POLY-PRO.

FT DOMAIN 1707 1712 POLY-ALA.
 FT 1952 1963 POLY-SER.
 FT DOMAIN 1014 1014 L -> V (IN REF. 2).
 FT CONFLICT 1508 1509 PP -> SS (IN REF. 2).
 FT CONFLICT 1561 1561 W -> R (IN REF. 2).
 FT CONFLICT 1567 1567 O -> H (IN REF. 2).
 FT 2440 AA: 270263 MW: 604AD7964D00EDB8 CRC64;
 SQ SEQUENCE
 Query Match 2.5%; Score 356.5; DB 1; Length 2440;
 Best Local Similarity 17.6%; Pred. No. 1.2e-05;
 Matches 449; Conservative 354; Mismatches 889; Indels 865; Gaps 108;
 QY 270 SAVYH--LECVKPLLEVPBDEMQ---CEVCYAHKPGVTCVAETIQNKRP----- 316
 DB 97 SPVDHDSLESKRPRLEQVDSHFQVSAVPLVPHLPGLRASADAKKDPAFGKHEAP 156
 QY 317 ---IRHEPIGYDSRRKKYFNLNRLIEDTEENENKTIWYTSYQVALIEDCLDKDW 373
 DB 157 SSPISGQPCG-----DDQNASPSKL---SKEELIOSMDRVREIA 193
 QY 374 EAE---LCKILEEMREI-----HRHMDITEDLTNRKSGNSFLAA 412
 DB 194 KYEQQLKLKKKKQQLLEBAKAPPEPEKPPVPPVQKHRSIQIITYDENRK----- 248
 QY 413 ANELLLESIRAKKGDIDNVKSPETE-----KXNETENDSKA- 451
 DB 249 EAKRIEGL-GRVVELPLYNQPSDTKYVHENIKTNQVMKKLILFKRRNHARKOREKI 307
 QY 452 -----EKNREFEDQSLEKSDSDKTPDDDEQKSEVGDKSEKSNELSESG 500
 DB 308 CORYDQLEMEWEKKVRIENPRKAKESKT---REYEEKPPELRKQEDQERQVWG 363
 QY 501 -AGKASGSTRITRLRNPDLSKLQKSOVAAAHAEAKLFEKGEVLYVNSGGEISRL 559
 DB 364 QRAGLSAT-----IASHHEISEIIDGSE--QENNEKMRGL 400
 QY 560 STKEVIMKG-----NINNYFKLGQCKRYVYNOYSTNSFALNKHQHRDHKRRH 611
 DB 401 SVIPMMFAEQRRVFKYFINMNGL---MEDPMKYKDRQFMNWT-----DHEK-EI 447
 QY 612 LAKKPCLTAGEFKKNGSVGSKVLITSLRLIQLQENNIPSEFLPMWASHRAWIKA 671
 DB 448 FKDKFIQHPK-NGGLASTLERSVPCVLYLYTKNNEY-KALVRNRYGKRGR- 501
 QY 672 VQMSCKPREFALALALECAVVKPVVMIPIWREFLGHTRLRMTSIREKEKVKKKKKQ 731
 DB 502 NQGIARPSQ-----EKKVEKEKDEKAKETKEKE 529
 QY 732 EEE-----ETMOQAT----- 741
 DB 530 EEKKDEEKEDEKESKENTKEKDKIDGTAETEEREOATPRGRKANSOGRRKGRITRSM 589
 QY 742 -----WVKYTPVKHQVKKQGEERYVGYG---GWSM 771
 DB 590 TNEAAAAAATAATEPPPLPPEPPEISTEPETSRKTE--EEKEVAKKGLVHGHRWA 647
 QY 772 ISKTHVREVPKLPK-----NTVNYRKSLGKRNKNNDEMDSDRKCSRSPEKTK 823
 DB 648 AA-----IAKVVGRKSEAGCKNFYFNKR-----RHMD-NLQDQHHQKSRKRE- 692
 QY 824 IEPDSEKDEYKGSDAKAGADQNMEDISKITEKKQDVKELLDSQKPKKEPMEVYDDM 883
 DB 693 -----ERD-----VSQCESVASFVSAODEDIE--ASNEEENP----- 723
 QY 884 KTESHNQCESSQVDVNVNSEGPHLTSYKKKKTKSSKLDGLLRKQKPTLEKQKLEKI 943
 DB 724 -----EDSEVEAVKPS-----DSPENATSRGNTPAVE----- 752
 QY 944 KLEGIGIKGIGTSTNSKUL---SESPVITKAKEGQSGQSMROEQSPNANNQPEDLIQ 1000
 DB 753 -----LEPTETAPSPSPSLAVSTKPADESEVETQVNSISGTEKQMDVQDQEH- 803
 QY 1001 CSQSDSSVLRMSDPSHTNKKLYPKDRVLVDVSTRSPETKCPKQNSIEND-IEEKVSDLAS 1059

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:21 ; Search time 45.9589 Seconds
(without alignments)
5817.155 Million cell updates/sec

Title: US-09-698-295-10
Perfect score: 14333
Sequence: 1 MYSEEEEDDGAETQDS.....RLKGFKASRSHNNKLGSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3363	23.5	810	2 G01252	small GTP binding
2	871.5	6.1	1711	2 T21432	hypothetical prote
3	522	3.6	2447	2 T16870	hypothetical prote
4	494.5	3.5	5327	2 T13564	microtubule-associ
5	459	3.2	5170	2 T15348	hypothetical prote
6	427.5	3.0	3759	2 A35085	trithorax protein
7	419.5	2.9	2232	2 T34434	hypothetical prote
8	413.5	2.9	3507	2 T34513	hypothetical prote
9	412	2.9	2897	2 B48666	cell proliferation
10	404.5	2.8	452	2 T21435	hypothetical prote
11	404.5	2.8	3256	2 A48666	cell proliferation
12	404.5	2.8	3488	2 T34418	hypothetical prote
13	402	2.8	1630	2 A53577	hypothetical prote
14	401.5	2.8	1367	1 S48478	glucan 1,4-alpha-g
15	399.5	2.8	405	2 T21433	hypothetical prote
16	399.5	2.8	510	2 T21430	hypothetical prote
17	399.5	2.8	2441	2 D71623	erythrocyte membra
18	397	2.8	2346	2 T13829	Trp homolog - frui
19	395.5	2.8	2271	2 F90073	hypothetical prote
20	394	2.7	2938	2 T30249	cell proliferation
21	387.5	2.7	1229	2 T23697	hypothetical prote
22	382.5	2.7	1589	2 T13606	hypothetical prote
23	382	2.7	2453	2 S60254	nuclear receptor c
24	375.5	2.6	3924	2 S37431	ankyrin 2, neuroma
25	374.5	2.6	2116	2 A26655	myosin heavy chain
26	374	2.6	4377	2 A55575	ankyrin 3, long sp
27	373.5	2.6	2541	2 T29340	hypothetical prote
28	372	2.6	5105	2 T32650	hypothetical prote
29	367.5	2.6	3259	1 A56539	giantin - human

30	367	2.6	2215	2 T16871	hypothetical prote
31	366.5	2.6	3225	2 T52300	giantin - human
32	364	2.5	6642	2 T29757	protein UNC-89 - C
33	360	2.5	1875	2 S38173	myosin-like protei
34	358.5	2.5	3828	2 T13857	trithorax protein
35	356.5	2.5	2187	2 T30826	nascent polypeptid
36	355	2.5	1459	2 T32271	hypothetical prote
37	353	2.5	2035	2 A40718	host cell factor C
38	352.5	2.5	1871	2 D96796	probable heat choc
39	351.5	2.5	3570	2 T45025	mucin MUC5B, trach
40	350.5	2.4	3147	2 T18674	hypothetical prote
41	349	2.4	2481	2 D90011	FnB protein limpo
42	347.5	2.4	2364	2 A56577	microtubule-associ
43	347	2.4	1939	2 T18372	repeat organellar
44	347	2.4	2422	2 T12687	ALR protein homolo
45	345.5	2.4	1302	1 J66009	surface-located me

ALIGNMENTS

RESULT 1	
G01252	small GTP binding protein SEC4 homolog - human
C:Species: Homo sapiens (man)	
C:Date: 21-Dec-1996	#sequence_revision 06-Jun-1997 #text_change 29-Jan-1999
C:Accession: G01252	
R.Bowser, R.P. submitted to the EMBL Data Library, January 1994	
A:Reference number: G06441	
A:Accession: G01252	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-810 <ROW>	
A:Cross-references: EMBL:U05237; NID:g1276427; PID:g451848	
C:Genetics:	
A:Gene: FAC1	
Query Match	23.5%; Score 3363; DB 2; Length 810;
Best local Similarity	82.6%; Pred. No. 3.1e-118;•
Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2;	
QY	1 MYSEEEEDDGAETQDSDEDEDEDEDDDDSDYPEEMEDDDDACTESSFNSHT 60
DB	1 MYSEEEEDDGAETQDSDEDEDEDEDEDDDDSDYPEEMEDDDDACTESSFNSHT 60
QY	61 YSSTPGRRKRVRRPSPRIEKKDIPLEPRKSSSDLMVNEHIMVATYEVLNRGTV 120
DB	61 YSSTPGRRKRVRRPSPRIEKKDIPLEPRKSSSDLMVNEHIMVATYEVLNRGTV 120
QY	121 LRISPRFEEDGCAIVSOECSTLMAEMHVVTLKAVLRSEDTSTTGGPADLKDSVNSTLY 180
DB	121 LRISPRFEEDGCAIVSOECSTLMAEMHVVTLKAVLRSEDTSTTGGPADLKDSVNSTLY 180
QY	181 FIDGMPTEVLRVYCESDKYNNHVLRYQEAEDYRYGFEVKNIKVLOFLVDOFLTTNIARE 240
DB	181 FIDGMPTEVLRVYCESDKYNNHVLRYQEAEDYRYGFEVKNIKVLOFLVDOFLTTNIARE 240
QY	241 ELMSGVYQYDHCRCVCHKGLDLCSTGSAVYHLEVCVKKPRLVEVRDEQCEVCVANKV 300
DB	241 ELMSGVYQYDHCRCVCHKGLDLCSTGSAVYHLEVCVKKPRLVEVRDEQCEVCVANKV 300
QY	301 PGVTDCAETQKKNRYRHRPIGDRSRKRYWFLNRLLIEEDTENNEKKIYYTSKVQ 360
DB	301 PGVTDCAETQKKNRYRHRPIGDRSRKRYWFLNRLLIEEDTENNEKKIYYTSKVQ 360
QY	361 LAELIDLDLDYWEALCKLLEEMREIHRHMDITTEDLTNKAAGSNKSFLLAANEELIES 420
DB	361 LAELIDLDLDYWEALCKLLEEMREIHRHMDITTEDLTNKAAGSNKSFLLAANEELIES 420
QY	421 IRAKGDIDNVKSPETEKKNETENDSKAEKNREEFQDSLEKSDDKTPDDDEQCK 480
DB	421 IRAKGDIDNVKSPETEKKNETENDSKAEKNREEFQDSLEKSDDKTPDDDEQCK 480


```

QY 1368 TKESTDTTTPSASCPESSNVQVEDMEIETSEVKKVTSPTTSEESNLNDPIDENG 1427
Db 1002 ----- 1001
QY 1428 LPINKENNGESKRKTVTEVTMTSTVATESKTVIKVEKDKQTVSTENCASSTYT 1487
Db 1002 -----SGVAGEEK----- 1009
QY 1488 TTTTIVTKLSTFGSGVDIISVKEOSKTVVTTVTDLSITTTGGLVTSMTVSKESYTD 1547
Db 1010 ----- 1009
QY 1548 KYKLMKESPPKTRGCTALPSRKVEVTKSTKSIYLPNDLKLARKGIREV--PYEN 1605
Db 1010 -----AMPIDELQTFSTKRGKSIYVLOKRIILFQIMGSGCCQVYMPGES 1054
QY 1606 YNAKRALDIPYSPRPRTGIMRYRLQTVKSLAGVSLMRLMLASLMDMAAKPPGG 1665
Db 1055 AGIKSMLLIMPYAPRPRTDLCWKQTLNARSLHAVALQKLIWSSIKNEFD--PDDT 1111
QY 1666 GSTRTESET--EITTEILKRDVPGIRFEYCIRKIIICPI--GVETPKETPTPOK 1722
Db 1112 HPRRRAVIDTSPDERRRRIIRHKEMRPYQYERYEMEITLPIYDEPEDESMLSRNG 1171
QY 1723 -----LRSSALPKRPET-----PKQT-----GPVILETVAAEELE 1754
Db 1172 GSSEFHSRSSARKKRPRHEFLSKFGVYPKSKNAFSLDRNRATAIRREWVDGVLTK 1231
QY 1755 LMF-----IRAFARVEKERQAV-----EQAKKRLDQOKTVIATSTTSPSS 1800
Db 1232 VEFIKDYWKIRAEKTKARKLEATRKAKAKADEDERRIQOQOVARIPV--PMHS- 1289
QY 1801 TSTISPAQKVMYPIGVS--TTGTMVLTGKVSPTATFQON-----KNFHOTF 1849
Db 1290 ----LISEKNNV--PYIGSQOQRRPNNGEFLKYNSSSVSQALGAYATPRPGHQP 1345
QY 1850 ATWVGQSGNSGVYQVQKVLGIIPSTGTSGQTFSPQRTATVYIRPNTSGSGGTTSN 1909
Db 1346 PNIIRQAGVNO-----LPKKPTTSPFNFS--RP--VATIPPTQLRAAGADG- 1390
QY 1910 SQVITPQIRPG-----MTVITPRLQOSTLKAIIITPVMQPGARQOVMQITRGO---- 1961
Db 1391 --VVRVMMTPGKSTVTNNTSTPYQ-----ALNRQOYQDQOQOQPAVRRLNGYHFM 1443
QY 1962 -----PVSTAVSAPNTVSSTPGOKSLTSAT----- 1986
Db 1444 GTYRGGGRPSVQNOHRLPQNRALORPGEESTEMKRYTEALIPNDGDGEOPVILRYD 1503
QY 1987 STSNISASQPRRPOGQVKTMAQLTQLTQHGSGNOGLTVVYIOGQGTGQLD---- 2042
Db 1504 PYSNFDQAQRAQOQHPSRPVYSTPACMIRTTQPGVKH--NVILMKASDGTQKMLKPG 1561
QY 2043 IPQGVTVLGRGQQLMAAMPNCTVORFLPTPLATTATTAATTTTSTTAAGTGEOROS 2102
Db 1562 FRPGTVI--STGQVVVYRQPTAVQOQOLYT-----ATPGTVVVRIPNANGAPRODH 1613
QY 2103 KL-----SPQOYHOK--TLPPAQ-----SSSVGPAKAQPOQAQPSARP 2140
Db 1614 QVNRKVVQASGRPRAMEYMDQGTTPRGQOQVRYVLOGGNSGTTPVNPBKVSSKGRPGGLT 1673
QY 2141 QPOTPOSPAQREPVOTQREVOV--QTIVSSHVPSEAO 2175
Db 1674 MQWVQOQOQHNPANHYMDRDTATGFAVSTTEQVYDEQ 1711

```

RESULT 3
 116870
 hypothetical protein t13h2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
 C:Accession: T16870
 R:WU, X.
 submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T13H2.
 A:Reference number: Z18593
 A:Accession: T16870
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2447 <WU>
 A:Cross-references: EMBL:U39653; NID:g1049397; PID:g1049400; PIDN:AA52494.1; GSPDB:G
 A:Experimental source: strain Bristol NZ; clone T13H2
 C:Genetics:
 A:Gene: CESP:T13H2.3
 A:Map position: X
 A:Introns: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1
 C:Superfamily: RING finger homology
 F:158-207/Domain: RING finger homology <RRN>

Query Match
 Best local similarity 3.6%; Score 522; DB 2; Length 2447;
 Matches 502; Conservative 413; Mismatches 1033; Indels 760; Gaps 114;

```

QY 9 EDGDAEETDSEDEDEDEDDSDYPEMEDDDSDASYCTESS----- 54
Db 2 DQSPGSTSKSARDKAEKNEENTSDDSSDEVASASEKSESPRSKKKVTIRVIPVR 61
QY 55 -----FRSHSTTSPGRRKRVNRPSPILEKDIPLPEPKSSBDLAVPNEHIMNY 108
Db 62 PRTDKGHRVNLSESGNESETKLYQAKEGIPSKKRPKELIPTTSEQYDDLEVLDMN- 119
QY 109 AIYEVLNFGTVLRLSPREFDFCALVSOECTLAEMHVVLLKAVLREDTSNFTFG 168
Db 120 -----PARNEG-----RELTLNA----- 132
QY 169 ADLGSVNSTLYPIDGMTPEVRLVYCESDPKEHNVLPYQEAEDYRGPYENKIKVQL 228
Db 133 --YDAVRKNYVNLPG-----KSYCEAD-----LQKV 156
QY 229 VDFLFTNARELMBGVQYDDH--CRVCHKLGDLCCETCSANYHLECVAPLEVP 266
Db 157 IGSF--SCDVCQELQGSIMTKKCGHRCDCILVAFMRSGNTPRTGROMLKSRELQDP 215
QY 287 E--DEMOCCEVNAKVPQVDCVAEIOKN-----KPYIRHPIGVDRSRRKWFNLNR 337
Db 216 RFQDLIYQVNESRSIVG--RMAENREHEKDYFGRRKTYEG--GSDMKRRKGIQPN 270
QY 338 LIIEEDTENENEKRI--WYSTKQVLAELDLCDKDYWEALCKILEMREIHRNDITE 396
Db 271 LKAPRLKSGRKKIRHNES-----DED--GSVRKVMSSKGA-----PKE 310
QY 397 DLTNKAQSGNSKFLAANEILSIRAKKDID--NVSPETEKKDNTENDSKDAEK 453
Db 311 DDTNYLENDEKGTSVAAKEVLE-----EGEMDFPIEIKSSDEQDLDDEESMLDSD 365
QY 454 NREFEFQSL-----EKSDDKTRPDDPRQEGKSEVDQFKSEKSGELES 498
Db 366 EISDNDEVSKPSCSTKKTNNRSRDSSESDNSRDNELQKKKKMKKKNVPKTDGSDVN 425
QY 499 PGAGKASG--STRITTYLR-----NPDSKLS--QLKSQVAAANAEANKLFEKREVL 549
Db 426 ESRDEDAASGEVATKILKEKKKPCGRRKKKFAPELLEGIIPRSEDSLSDEBRD 485
QY 550 VNS-----QGEISRLSTKREVIKNGINNY-----FLGDBGKRYVYHNOTSTN 553
Db 486 ADYAEVFGQKEFNDRDRDGHPRKDKLYNDFMIDMNHQVRFKKGELIHYISD--DSN 543
QY 594 SFALNKQHREDDKRNHLNKKFLPAGFGKNGSVHSGKVLITSLRLITIQLENNIP 653
Db 544 S-----EHESDEADRESSI-----DSHEKEI----- 566
QY 654 SFLHPMASHANWIKAYOMCSKPREFALALAILCAVVPVMLPIWREFL-----GHR 709
Db 567 SKFL-----SHR-----QRLNPPTS-----VDDQOQVITVKKQVKSALTSKGETS 609
QY 710 LHRMTIEREKKRVKKKEKQEEBETMQOATVWKTTPVKKQVKKQKGEETVYTGCG 769

```


A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025392
 A:Introns: 24/1; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A:Note: EG:49E4.1
 C:Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 3.5%; Score 494.5; DB 2; Length 5327;
 Best Local Similarity 17.7%; Pred. No. 2.6e-10;
 Matches 527; Conservative 478; Mismatches 1189; Indels 791; Gaps 119;

QY 1 MYSEEEEDDGAETOD--SEDDEDEMEEDDDSDYFEE-----MEDDDDASC 50
 DB 1751 IASEASRPASVAESVKDEAEKSEESRESVAEKSPLPSKASRPASVAESIKDEAEKS 1810
 QY 51 TSSFRSHSTYSTGRRKPRVHRPRSPLEKEDIPLEFPFSSSEDL-----VFN 101
 DB 1811 KEES--RRESVAEKSPLPK--EASRPASVAESIKD-----EAEKSEESRESVAEKSPLPS 1864
 QY 102 EHMNVIAIYEVLRNFGTVLRSPREFDFCALVSOEQCTLMAEMHVLLKAVLREEDT 161
 DB 1865 KASRPASVAESIKD-----EAEKSEESRESVAEKSPLPSKASRP 1904
 QY 162 SNTTGPADLAKSVNSTLYFIDGMTPPEVLRVYCESDKYHHVLPYQEAEDYPRGVEKN 221
 DB 1905 SR---PASVASIKDE-----EAEKSEESRESVAEKSPLPSKASRP 1944
 QY 222 IKVLOFLVDQFLTINARELMSEGVIOYDHCVRCHKGLDLCCTGCAVYHLECVKRP 281
 DB 1945 ASVASIKDE---EAEKSEESRESVAE-----KSP 1972
 QY 282 LEEVDEDEMOCEVCAHAKPGVTDCAETQKPKYIRHEPIGYDRSRKRYWFLNRLLIE 341
 DB 1973 L---PSKEASRPASVAESIKD-----EAEKSEESRESVAE--EKSPLPSKASRPASVA 2022
 QY 342 EPTENENNEKIYVSTKVQALIDCLDKDYEAELCKLLEMEBEITHHMDITEDLTNK 401
 DB 2023 ESIKDEAEKS--KEESRESVAEKSPLPSKE--ASRPASVAESIKDEAEK---SKESRR 2075
 QY 402 AAGSNKSF-----AANEIELESIRAKKGDIDNVKSPETEEDKDNTEENDS--KDEAKNRE 456
 DB 2076 ESEAERKSPLEKASRPASVAESV-----DEADSKESRESVAESKAS 2123
 QY 457 EEDDS-----LEKSDDKTPDDPEOGKSEVDFSEKSNGLSESPGAKGA 505
 DB 2124 IKGDSPLEKVERPSVAESVKDDPVKSKEPRESVAESVADASARD--QSPLESKA 2181
 QY 506 SGSTIITIRLRNPDKLSQKQVAAAHAHANKLFKEKEVYL-----V 549
 DB 2182 SPRESVDSVKDEAEKQESRRESKTESVTPRAKDDKSPKELQPVSMETITREDADQPM 2241
 QY 550 VASGEISRLS-----TKREVIMKGNINNVFKLGQEGKRYVHNQYST 592
 DB 2242 KPSQAESRRESIAESIKASSPDEKSPASKASRPASVAESIK-----YDL 2288
 QY 593 NSFALINKHQHREDHDKRRHLAKHFKCLTPAGEFKWNGSVHGSVLTITSLTITOLENNI 652
 DB 2289 DKPQITIKDDKSTEHRSRELEDKSAVTS-----EKSV--SRPLSAVSDHEAAVATEDDA 2340
 QY 653 PSSF-----LHNMW-----ASHRANVIKAVOMCS----- 676
 DB 2341 KSSISPKDKSRGFVAETVSSPIEBATMEFSKIEVEVKSLSLALSGGSGKRLQTDSSPV 2400
 QY 677 --KREFALALALECAVAVMLPIWREFLHTR-----LHRMTSI 716
 DB 2401 DVABEDFSAHVAASV--TYTPTILTKPAELAOJGAATVSSPLDEALRTBPAPHISHADSP 2459
 QY 717 EREKEKVKKKEKQEEETMOQATWVKYTFPVKHQVOMKQGEYRVYGYGWSWISKTH 776
 DB 2460 AECASEEIIASODKSPQVLKESRPAMVAESKDDAQL--KSSVEDLR-----SPVASRE 2511
 QY 777 VYRFPYKPLPGNT-----NNYKSKSLEGTAKNNMD----- 804

DB 2512 ISR--PASGETASSPIEFEPAPDFAEEQAEKAVLPLTELKGNLFTLSSPVDVAHASVQ 2569
 QY 805 -ENMDESDRKRCSRS-----PKKIKIEPDEKDE-----VGSDPAK----- 840
 DB 2570 PAELSLVLDLEKTAASSPIDAPKSLIGSPAEERRESVAESAKDAEVEKSKDASRPSPV 2629
 QY 841 ---GADONEMDIS---KITEKKDQDKELDS-----DSDKPCKEENPEVND 881
 DB 2630 ESTKADSTKGDISPSPESVLEGPDKDVEKSESRPPSVASITGSDTKV--SRPASVE 2688
 QY 882 DKTESHVACQSSQVDVYVNVSEGFHLRTSYKKTKTKSSKLDDGLERIKQIFLLEKQRL 941
 DB 2689 SVKDEH--DKASRRRESIAKVES--VIDAGKSDSKSSQD---SQDKKSTLASKASR 2741
 QY 942 KIKLEGGIGICKTSPNSKNL-----SESVITK--AKEGQSDSMHROGSPNA 989
 DB 2742 RESVYSSKDDAEKSSRRESVYASGEPPRESKSLDKDTSRPSVSEYTADEKSE 2801
 QY 990 NNDQPEDL---IQGCSQSDSVLRMSDPSTTNKLYPKDRLVDVYSIRPETKCPKRONSI 1046
 DB 2802 QOSRRESVAESYKADTKKQKSOEASRPSSVDLL--KDDDEKQESRQSIGSHKAMST 2859
 QY 1047 ENDIEKVSDDLASROEPTK-----SKTKGDFIIDSKLASADDIGTLCKNNK 1096
 DB 2860 MGD--ESPMDKADKSKEPRESVAESIKHENTKDESPLAGSRORSVAESIKSDITKGEK 2917
 QY 1097 -PLIOBE---SPTIYSSK-----SALHSVSKSTNDRAPPLSRAMDF--EGKLGK 1142
 DB 2918 SPLPSKEVRPESVSGIKDEKASRESVAESVPESSKDATSAPPSKESHPRESVLS 2977
 QY 1143 DSESNSTLSSNDTVSIODSSEED---MIVONSNEISQOFTREODVLEPLKCELV 1199
 DB 2978 LKDEGD--KTSRRVSVADSIKDEKSLVSOEASRESE-----ASLKDAAP 3024
 QY 1200 GESTGNCEDRLPYKGTAEANGKRPQOQKLEBRPVNKCQOIKLNTTDKKNNEKRESEK 1259
 DB 3025 SQETSRES--VTESVKQKSPVASK--ASRPASVAEN---AKOSADSKQDREPSLQ 3077
 QY 1260 GORTSTFOJNGKDNKPKIYLGCELKEISESRVSGNVNPKVNIKITPENDIKSLTYK 1319
 DB 3078 SKAGSI---KDEKSPASKDEAEKSEESRES-----VAEQPP-----LYS 3116
 QY 1320 ESAIRPFGVDVIMEDFNERNSETKSHLLS---SDAEGNYRSDLETLPSYKESDST 1374
 DB 3117 KEYSRPASVAESYKD---EAEKSEESPLMSKASRPASVAGSVKDEAEKSEESRESV 3173
 QY 1375 QTTTPASCPESNVQVDMETISEYVK-----VTSPTTSEENLNSDPTIDENG 1427
 DB 3174 AEKSPLEKASRPASVAESVDEADKSEESRESGAERKSPASKASR----- 3223
 QY 1428 LPIKNENYNGESK-----RKTVITEVTMTSTVATEKTYIKVKGKQTYVSTENC 1481
 DB 3224 -PASVAESIKDEAEKSEESRESVAEKSPLPSKASRPSTSAKSKDAE---KSEES 3279
 QY 1482 AKSTVTTTTTYTK--LSPSTGGSVDIISVKQSKTAVTTVTYDLSLTGTLVTSMTVS 1540
 DB 3280 SRDSVAEKSPASKASRPASVAE---SVQDEAEK-----S 3312
 QY 1541 KEYSTROKV--KLMKSRPKRTSGTALPBYRKRYKSTKKSIFVLPNDLKLARKGGIR 1599
 DB 3313 KEESRESVAESKPLAYKASRPASVAESIKDEAEKSEES-----RESVAE 3360
 QY 1600 EYVYFYNNAKPALDIMPYSPRPTEGITRWRYRLQYVKSLAGSLMLRLMASLMDMDMA 1659
 DB 3361 KSPASKAS-----RPT--SVASVADAEKSEESR-----DSVAE 3397
 QY 1660 KVPPEGGSTRTEETSETTEITTEITIKRRDVG---PGAIRFYCIKRIICPIGVETPKETP 1716
 DB 3398 KSPASKASRPASVAESVQDEAEKSEESRESVAEKSPLASKASRPASVAESYKDDA 3457
 QY 1717 TPORKGLRSALRPKRPETPKQTG--PVILETVABELELMEIRAFAEVYEKQAVQ 1775
 DB 3458 EKSKESRESVAEKSPASKASRPASV-----AESV--KDEAEKSE 3499


```

QY 1776 QAKRLDQKPTVATSTSTSTSTSTSPAKQVWVAPISGVTQTKMLTKVGS-P 1834
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3500 ESRRESVAEKSPSPKESRPTVAESVKEAEKESRESRESV--AEKSLASKESRSP 3557
QY 1335 ATYTFQONKNHOFATFWWKOQOSNVVOYQKVLGIIPSSGTSCOTTSTFOPRTAV 1894
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3558 ASV-----ASVDEAEKESRESRESVAEKSPKESAS-----RPASVAE 3599
QY 1895 TIRPNTSGSGTTSNQTNGPQI-----RPGMTVIRFLQOSTGLKAIIRTPVMQPG 1948
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3600 SVKDEAKESKESRESRESVAEKSPKESASRP-TVAESVKDEADKESRR-----ESG 3654
QY 1949 APQGVMTQIRGQPVSTAVSAPMTVSSTPGOKSLTSTSTNIQSASQPPRQOGVKL 2008
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3655 AEKSPKESASRPTVAESVKDETEKESRESRESVTEKSPKESASRPTVAESVK- 3713
QY 2009 TMAQTLTQGHGNGCLTVYIQGGQTTQQLLIPQGVTVLPGPGQQLQMAAMPNGTVQ 2068
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3714 -----DEAEKESRESRESVAE 3730
QY 2069 RFLFTPLAT-TATYASTTTTSTTAGTGEOROSKLSPOMYHODTLPPAOSSVGP 2127
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3731 K-----SPLASKESRPAVASINDAEGTQOESRRESMPS--GKAESTIGDQSSLASKE 3785
QY 2128 KAOPQTAOPASARPQOTQOPSPA--QPEVQTOPEVOTQTTVSSHVSEAOPTAOSKQP 2185
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3786 TSRPDSVESVSKDETE-KPESGAIKQSVASRPE-----SVAVSADKESKPLH--SRPE 3836
QY 2186 VAAQSQSQSVQOSQPVKQVSPQSTRIR--PSTPSQLS-----PGQSOVQTTSTQPIPI 2238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3837 SVADKSPDAKESRSLSVAETASSPLEEGPRSIADSLPLNTLGEKRGKPLTSLPIDV 3896
QY 2239 QPHTSLQIPSGQGPQ-----SQPGV--QSSQTQTLSSGQTLNQVSV- 2276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3897 AEGDFLEVKAKESRPAVLSKPAEFSGPDTGHASTPVDASVYLEIEVVEQHTTSGVG 3956
QY 2277 ---SSPSRQLOIQQPOQVIAVPQLOQOVVLISQIQSVVAAIQAOQSGVPQ-QIKLQL 2332
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3957 ATGATATETDLDLLETKESEVT--KQSETTLEFETLSKVESKVEVLESSVKQVEEKVQT 4013
QY 2233 PI-----QIOQSSAVQTHQIQNVVYQVQASVQEQLO-RQOQLRDQOQKKQOQ 2379
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4014 SVQQAETTVTDSLQELTKKSEQLTEKSVLDTNINSVTNLSTAVETLEKKVQDVTETK 4073
QY 2380 IEIKREH-----TLQASNOSEIIOKO-----VAKKHNA 2407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4074 IEKATEHVSEHVTTTGESSTSTQSEKSSDLGTFSELRETHITTVGSPETTVYICERDEP 4133
QY 2408 VIEHLKO-----KKSMTPAE-----REENORMITCN--QVMKY---ILD 2441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4134 VLHDIEEEDHEHRRSPSPDVDAKAIIPQPMRPLSPREEVAKIVADVAKVLSKDKDID 4193
QY 2442 KI-DKEKQAAKRRKRRESEVEQKRSKQNTKLSALLFKHEQIRAILKKRALLDK--D 2497
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4194 ILPDFERQLEELKSTADTEESDKSTYDEKSLFI---SVKVEIESEKSSPDQKSGP 4248
QY 2498 LQIEVOEELKRLKIKKKDLMLQATVAAACPPVTVLPAKPAPPSPPSPPGVQHT 2557
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4249 ISTEEKDKIQSEKALQRLQGLASSRPEVA-----SQPESVP----- 4286
QY 2558 GLISTPLPYVASQKRRKRREKSSSKKKKKMISTTSKTKKDTK 2602
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4287 ---SPSQAASHHEKLEVELSESHKAEKSSRPESVASQVSEKDMK 4327

```

```

RESULT 5
T15348
Hypothetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15348
R:Gatlung, S.

```

```

submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PID:AA93447.1; CESP:B0
C:Genetics:
A:Gene: CESP:B0350.1
A:Introns: 48/1; 5039/3; 5116/3

Query Match
Best local similarity 16.9%; Pred. No. 5,3e-09;
Matches 577; Conservative 515; Mismatches 1157; Indels 1162; Gaps 142;

QY 4 EEEEEEDGAEETQDSE-----DDEEDMEEDDDSDVPEMEDDDDDASYCTESS 54
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1655 DETRSEAGDSHIEYETTKTTTVREFHGEQPEETEETDEVEELPPKILEEDNVSESSST 1714
QY 55 FRSH-----SYSSNPGRRKPRVHRPRSPILIEEDDIPLEPPKSEDLMPVNEHI 104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1715 SVSREVRPDPPEHIIETTTTTVTREYN-NEPEETDYDQDAAPISFSQHODDQASHD 1773
QY 105 MNVAIAYEVLNMGVLRSLSPREFEDFCAALVSQEOCTIAMEHVVLLKAVLREEDTSMT 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1774 QHD-----RSPVSE-----KSVKHTTETTTT 1796
QY 165 TEGPADLKDSVNSTLYFDICMTWPEVLRVYCS--DKKEYHNHLYQ--EAEDPYGCV 218
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1797 TVTTRQLYDD-----EASEIRGESPVATEEHEHVSSTKSDSESOH----- 1837
QY 219 ENKIKVLOEVDQFLTNTIAREELMSEGVLOYDDHCRVCHGLGD-----LCCE 267
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1838 -----VPSVIEETTTTTVTREPYDDQDLQREDHTQSEKRSSTIETETHEDSHLIKE 1891
QY 268 TC-----SAVYLCEVCPPLIEVP----- 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1892 TTTTTVTRREFYDEPENEKLDQSPSLSPSSHVESEIYVPSPAKQOEIQTRFPHED 1951
QY 287 -----EDMOCYCVAAHKVPGVTDCAVLEIQKKPRTYRHPRTIGYDRKRRKTYWFLNRL 338
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1952 SPAQYFHEHDEYE-----HOVPTQAPRLTEQOHQPESEGSDEGFGSKVLGPAKKAG 2005
QY 339 II-----EEDTENENKATIMYSTVQALALIDC----- 367
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2006 MAAGGVAAAPVAAVGAATAADALKKEDDEDEQERE-----SLIRERSDSHASEQ 2060
QY 368 --LDKDYWEAELCKIILEMREIEHRMD-----ITEDLTNKRAGSNKSFLAANEE 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2061 SOLTEKH-----KREESPVPSEKHNHDOSSALPQESVQPIEKESRTFNDE--SDF 2109
QY 417 ILESIRAKGDDIDNKSPE-----TEKDKNETENDSDAKENREPEFDQSLSDS 467
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2110 GVKSDHYTDDQSLKSPKESGEAFSQFTSEKEQORSBPISQKEDISQFQNESSPBDV 2169
QY 468 DDKTPDDDEQKSEYGDKSEKSNELSESPGAGGASGSTRITRLRNPNPSKLSQKLS 527
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2170 KSFQPHDEKPLERQSGT-----SSGYSPKSPGSGITGIDEKALSGVQEPDRPRENPAE 2225
QY 528 QQVAAAHAENKLFKEGK-----EVLVNSQGEISRLSTKKEVIMKGINNYFKLGQEGK 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2226 SHEKTQATSDENLFEQDKYAPASVPVSEDSNRYLETITTTTV-----TREHFEEDDS 2280
QY 583 YRYVHNOYS-----TNSPALKKHQHRED-----HDKRRHLAH 614
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2281 YVESQEYSSSGSPVPSEKSVDRVLETITTTTVTRREHFEADEIPTIYESSHDDQ----- 2335
QY 615 KCPCLPAGEFKNGSVHSGKVLITSLRLTLQNLNNIPSSPL--HNNMASHRANW--I 669
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2336 -----AASVPSEEDVH--EQIQTTTTTTTTTVK-EHVPDEIDSEHNNESKYSAGSPV 2387
QY 670 KAVOMCSKRREFALAIL-----ECAVVKPVVMLPIWREFLGHTRLHRMTSIEREEKK 723

```


Db 2388 PSEDESSRVETTTTTVTRHFEPEDDHS PVQ-----TOEVSASES 2430
QY 724 VKKEKKEOE-ETMOQATVWKYTFPVKHQVWKOKGEERYVTGCGMSLSTKHVYRFV 782
Db 2431 PVPSEKSEVERIETTTTTVTRHFEDEHILGQOSES-----2468
QY 783 KLPQNTNVRKRSLEGTCKNMENMDESDKRCRSRPPKIKIEPDEKDEKVGSDAKGA 842
Db 2469 -----DSQIPSEISITSENM-----RETSSPVQ-----SNRDE-----2497
QY 843 DONEMDISITEKKODVVELDSD-----SDKPCKEPMEV---DDDMKTESHVNOEES 895
Db 2498 ---EVLPAIAPYKOPTEXGRVDSHDAPASAPASAPISQAYKQESOEKHSLEOQORS 2554
QY 896 QVDVVA--VSEGFHLRTS--YKKTKSKLDGLERIKOFTLEKORLEKIKLEG---G1 949
Db 2555 SVSHESPAAOYFHNDSEHSDSPVSDRAPLLEQOHQESGESDQ-----EGFGSKY 2608
QY 950 KSIGKTSTNSKNLSESPVITRAKEGQCS--DSMRQOSPNANDOPEDLIQOCSQSDS 1007
Db 2609 LCFAKKAGVAGVAVAPV--ALAAVAKAAYDALKKDDDEDENDPQKLLP--KSPERQ 2665
QY 1008 VLKMSPTTKLYPKVDLVDSIR---SPETKCPKONSLENDIEEKVSDLASRQOE 1063
Db 2666 VL-----VMPVEPSEISEIELEYTSPSETEKSESQCYTE-----2701
QY 1064 PTKSKTKGNDFIID-----SKLASADDICLICK-----NKKPLIOESDITVSSSKSA 1113
Db 2702 -IVRTTIVREIYIDDOSTVRSRSEHDISEQYAPESPEODPYVEKTTVITROYHDE 2760
QY 1114 LHSVVKSTINDRAPPLSRAMFEGKLGCDSESN--STLENSSDTVSIODSSEEDMIVQ-- 1170
Db 2761 PPOEIEQOTIPEBVYLVREV--YESPEGDEPQHYIETKTTTITKEVHVAVEDVQIS 2818
QY 1171 -NSNEISF-QERTROQ-DVEVLEPLKCELVSESTGNCEDRLPVKGTANGKRPQOQK 1227
Db 2819 VHSSTVSSEKQPLADQLEPQESTTATVTREREVEEVPVPGSEADDE--SHAPK 2876
QY 1228 LBERPVNKCSDQIKLKTNDKNNRESEKKGQRTSTFOINKDKMPKITYLGECLKEI 1287
Db 2877 YMETTT-----TTTVTREREVESEDEHQSOV-----QRDSPPA--PSEDSVKHY 2920
QY 1288 SE-----SRVSGNVEBKVNINKIIPEND-----IKSLTVKESAIRPFTNGVIMEDFNE 1338
Db 2921 IKTITTTVTEREYBE--DSHSPVSEDDVHGFKVTTTTTT-----VTHHFEP 2969
QY 1339 RUSSEKSHLLSSDAEGYRDSLETLPTKESDST---QTTTPASCPESNSVNOVEDM 1395
Db 2970 EDPSPDEHVESERYASGS-----PVPSEDSREIETTTTTTVTRHFELEDDEH 3021
QY 1396 EIEETSEVKVTSSPITSEE-----ESNLSN-----DFIDENGLP-----I 1430
Db 3022 VVESQOEY-SAGSGPVPSEKSEVERIETTTTTVTRHFEHEDDIPITVETSHDDPPASSV 3080
QY 1431 NKNENANGESKRRVITVETVMTSTVATE-----SKTVIKVERGDKQIVVSTENC 1481
Db 3081 PSEEDVHGQ-----IQTTTITTVTRHVVDDEIDSGRMDLEKYSSESVPSEDS 3133
QY 1482 ASSTVTTTTTYYTK-----LSTPSTGGSVDIISYKEDSKIVVTTTTVDSLT 1527
Db 3134 SVRIETTTTTVTRHFEDEDDSHVESQOESASGS-PVPSEKSEVERIETTTTT-----3188
QY 1528 TTGGTLVTSMYKSEKSTPDKVLMKFSRPKPTRSGTALPSYKFEVTKSKKSIFFLPPND 1587
Db 3189 -----TYTRHFTDEEYI-----PSBSRT-----SHDDGITDOHVPSQSPVSE 3228
QY 1588 DKKLARKGIREVPY-----FYNAPALDIMVPSRPFTGJ-----1626
Db 3229 EDDHTHEQIINDDPIDEQIVESHEKSESST-----PSEQOSTHYIETVMTSPINSEKX 3283
QY 1627 -----TMYR-----RLQTVKSL-----1638

Db 3284 DPEVEKDVESADEIDSDSTAQYKSESPVOTEKSIILLAKOQOESDSESGFGSKVLG 3343
QY 1639 ---AGV-----SLMLRLMASLRMD-----DMAKVPPOGG-----1666
Db 3344 FAKKAGVAGVAAVYALAAVAKAAYDALKKDDDEDEDEKBPILIGKHKODPI SOD 3403
QY 1667 ---STRFETSEITETTE-----1681
Db 3404 EPSAVSETOPEPATREYEPREEDKVIITDSASVQDEPKIVPVDSTPEHNSNDREEFE 3463
QY 1682 -IKRRDVGPGYGRFEXCIKTKICPIGVETPK-----PTPQORGLRSALRPKR-- 1732
Db 3464 STVKSE--GPY-----IVESTDYAQTSAEPRISSPVHSDAGDASFFRSPESVTG 3511
QY 1733 -----PETPKOTGPVLIETVWAELEL-----WETRAFEREKEKAQAVQOAK 1778
Db 3512 EDEKNAIPETSEIDAYVIDS-----EYEFNNNDQORISSPAISDEDEDAVIDSEFY 3567
QY 1779 KRLQOK---PTV-----IATSTSP---TSSITSTI-----SPA-----1807
Db 3568 RHQOEONNEEDPSESIVESGEYISSGHGSPREFEDSTTTVLTNVHHEPAALPEPEVEDELE 3627
QY 1808 -QKVMV-----APISGVTTGTMVLTLY-----GSPATVTPQOKNFHOTATWY 1853
Db 3628 QERSIIESEYKTSPLPPTSVTVHEHVEPAELHKYRTSPRTVTVVSEHLDNRQEPY 3687
QY 1854 KQG-----OSNGVVOOVOKVLGIIPS-----STGT 1879
Db 3688 VESEETTRASPPLPERPESGSPLEREDDSHVISHETSSPVSEDSVKVIEKTT 3747
QY 1880 SQOFTSFQPR-----TATVTRPNTSGSGTTSNSQVITGPOIRD 1920
Db 3748 TVTVEREYEPEDSHSPVSEDDVHGFKVTTTTTTVTHHEFEDEHDSDEHVESRYAS 3807
QY 1921 GMTVITPLQOST--LGKALIRPVAVQCARQOQWQIIRGCPVSTAASAPRTVSTG 1978
Db 3808 GSPV---PSEENSNRYETTTTTTVTRHFEPEDDHVESQOESYA-----SGSPVPS 3858
QY 1979 OKSL-----TSATST-----SNIOSSASQPPRP--OGVKLMAQ 2013
Db 3859 EKSEVAVIETTTTTTVTRHFEHEDLPITVESHHDQOASVPSSEEDVHGQIQ--TTTTT 3917
QY 2014 TOLJOGHGNQGLTVVIOGGQOTTQOLIPQGVTVLPGGQOLMOAMPNGTVOREFL-- 2071
Db 3918 TVTTRH-----VVPBEIDSGRMDLEKYSSESVPVEEDSSRVLETTTTTTFIRE 3969
QY 2072 -FTP-----LATTTATAS-----TTTTVSTTAAGTGQORQSKLSPOM 2108
Db 3970 HFEPEDDSHVSGOERYASGSVPSEKSEVERIETTTTTTVTRHFEHEDHA-----4023
QY 2109 OVHODTLPRAOSSVGPAPKACQOTAPQPSARPOQIOPQPSAPPEVQOTOPEVO--TOTTV 2166
Db 4024 QLHD--TVSEKTSISSEPEQOTLEADITEETHQORSAEDENOP--ISTEKPVHMETTS 4080
QY 2167 SSHVPEAOPTH-----AOSKPOVAASQPOSNOVG--QSPV-----2202
Db 4081 TSHTASHVELEHDDDESGEGFGSKVLGFAKKAGVAGVAVYALAAVAKAAYDA 4140
QY 2203 -----RVQSPQ-----TRIRPSTQSLSPOQSO---VQIT 2231
Db 4141 LKKDDDEATYDLDRHGSYSAELQRPQOQVSERODESESGSFVHGDSOHLRVOQST 4200
QY 2232 -----TSOPIPIQPHSLQIPSOQPOQPOVOSSQOTLSSQOTLNOVSSPSRPOQ 2285
Db 4201 EAAFKLSSEIDIAQ--TAEKLVSEVAPLDRILISEQIRASIESLNR--NSPVKPS 4255
QY 2286 IQQPOPOVIAVPOLO-----QOVYLSQIOSQVAVQ 2316
Db 4256 VED-----LASEALQVNTKISFHADESDDKHDGNDGEMKYVDNRGVEVLEERSQTLTD 4310
QY 2317 I--QAOQSGVPOQIKLOLPIQIOQSSAVOTHQIONVYVQAASVQOLOLRVOQLRPOQK 2374
Db 4311 VIOAEGDATTHIMTOAEYSPRSKFLKQESCOEISNEPEVDYSDLO--EKLNILAGE 4368

QY	2375	KKQOQIEIKREHTLOASNOSELQKQYVMKHNAYIEHLKQKSMTPARREKOR-----	2428
Db	4369	KNNLHAIIVEEBSSSASOLOVTHES--DOEHLHEOHQOEAAQKODEKEEAEYTA	4425
QY	2429	MIVCNOVMKYLIDKIDRKEK-----	2456
Db	4426	TLIVDVYIQOYVOEINEEDDKTYMTSTSYLTATEKDOEYDTCVTSODDYYESQGWTS	4485
QY	2457	EESVEOKSKONATKIS---ALFEKKEOLRAEILKRRALL--DKDLQIEVOEELERD	2509
Db	4486	QDESYTTATSQAPRSLSDSADIELTARHDOERQETSTQALISPVSDRHFTYQOOFEMP	4545
QY	2510	LKIKKEKOLMO-LAQATAVAA-----PCPVPVYL	2538
Db	4546	VIRAFDPDDFMOTTRSTPVALQVTEIEEDESDDKLPISPSGILLPKHDPGRISVPP	4605
QY	2539	P-----APAPPPSPPPPPV-----QHGGLL--STPT	2564
Db	4606	PKRSDGTMOKEGDHFVYREEDTIAEPPPPPOPAPEOLAEEYATKESEHOOTIEGDVQO	4665
QY	2565	LPVASOKRRKEEK-----DSSSKKKKKMIS--TTSKETKKDI	2601
Db	4666	EEETEMRQETERIHSLAMEASSDLCNSESSESRYSKQSLSDISSKSESHADT	4716

RESULT 6

C:TriThorax protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
 C:Accession: A35085; A38240
 R:Mazo, A.M.; Huang, D.H.; Mozer, B.A.; David, I.B.
 Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
 A:Title: The triThorax gene, a trans-acting regulator of the bithorax complex in *Drosophila*
 A:Reference number: A35085; MUID:90192757; PMID:2107543
 A:Accession: A35085
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-3759 <MAZ>
 R:Mazo, A.M.
 Submitted to GenBank, January 1990
 A:Reference number: A38240
 A:Accession: A38240
 A:Molecule type: mRNA
 A:Residues: 1-2361,'Y',2363-2397,'N',2399-2405,'N',2407-2411,'N',2413-3759 <MAZ>
 A:Cross-references: GB:M31617; NID:g158818
 C:Genetics:
 A:Gene: FlyBase:trx
 A:Cross-references: FlyBase:FBgn0003862
 C:Superfamily: *Drosophila* triThorax protein
 C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match	3.0%;	Score 427.5;	DB 2;	Length 3759;
Best Local Similarity	18.4%;	Pred. No. 5.3e-08;		
Matches 609;	Conservative 407;	Mismatches 113;	Indels 1173;	Gaps 149;

```

QY 5 EEEEDDGAETOSSED---EEDMEEDDDSDDYEEEMDDDDDAASYCTESSFSRSH 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 DEDDEGVTFRRNDSPEEDONNAEDDEDDDDDDDEADEADQONEDNDNEASSESAETEK 597
                                     -TPGR--RKPRHRESPDLEEKDIPLEPKSSSEDLWPN 101
QY 61 YYS-----TPGR--RKPRHRESPDLEEKDIPLEPKSSSEDLWPN 101
                                     : : : : : : : : : : : : : : : : : : : : :
Db 598 AGADERPDEKQIWMDSHFVLPKRSTSSRIKRNKLLBEGAL-----STKKPLSLGD 651
                                     : : : : : : : : : : : : : : : : : : : : :
QY 102 EHIMNVTAI-----YEVLRNFGT-----VLRISPPRE 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 SKGRNVEGTSSSSAGSTASTFSASTNKLTKEFTFNFGTLKPNSSAAGIVFLRQPIRFQ 711
                                     : : : : : : : : : : : : : : : : : : : : :
QY 130 DFCAALVSQECITLMAEMHYVLKAVLRREDTSNTTTPGPAIDKDSVNSTLYFDGHTWPE 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 -----ADNODATTAARKACOTPSAIPKRNANSLATISSFGLASTNSTNY-----TPPS 760

```

QY	190	VLRVYCE----	SDKRYHVLHPQOEEDPYRYPENKI--	KVLOFIVDOQFTTIAAEELMS	244
Db	761	ACST--CSAVYSKEV-----	TQAKRYVAVADYCKRFPSFKMTKKTSANSTAN--TS	810	
QY	245	EGVIOYDHCRCVCHKLDDL-----	CFTCSAVYHLECVAPRPEEYPEDEMOCEVC--	295	
Db	811	SGSOQY-----	LQCKNGESSPCSISAKSOLKNFKFYKD-----	852	
QY	296	-----VAHR-----	VEGYT--DCYAELOQKKKPIIRHEPIGYDBSRKKYWFLLNR	337	
Db	853	KCMISFOLPAHRSRLSAILPPMRGEAAAREKSAELL--	SPGSLNF-----	900	
QY	338	LIIEEDFENEKEKIWYSTRKVOLAELIDCLDKDYEA--	ELOKILIEEMREIEHRMDIT	395	
Db	901	-----TSTASSSPSV--VASTSVC-----	WKSSGDSTSLTSTIKPPLAENNVY	943	
QY	396	EDLTNRKAGS---NKSEFL-----	AAANEELIESIRAK--KGDIDVNSPEBTEKD	441	
Db	944	FSGPPLRLPALLENPLFLKISNADQKILAAEAELISPLTKRNSKQEKREKVSQSEKLL	1003		
QY	442	NETENDSKD---AKNNEPEEDOSLEKDSODKTPDDDDPBOGSEVGDKEKSENGELES	498		
Db	1004	SPTQAGTKKSGAAGAOUEEYOPEQOEAPQSTTTPOPSASNGASH-----	GVPOA	1052	
QY	499	PGAK--GASGSTRIITLRLNRPDSLSQL-----	KSQOVA	531	
Db	1053	ELAGETNATGTLTKRQIDLPGRVRKHVCRSASIYGLPLATFEGDOQPEDAADMOEILA	1112		
QY	532	A-----	AAHEANKLREKEKE	546	
Db	1113	APVSAIMEPSPKPTHIVIDENDNCASCSTPSVGDSEKSPKSSSGAQAQEVKATATALKE	1172		
QY	547	VLVVNSOGELISRLSTKEKVEJMKGINNYFLGOEGKYRYIHN--	QYSTNSFLANKHQR--	603	
Db	1173	GTASAGGSSAKVYTRNAVAASNLIVAASKKQKNGDIATSSVYTOSSNOTQRTKTEHRQ	1232		
QY	604	-----EDHO-----	KRRHIA-----	HK	615
Db	1233	QRTILSIDFENWIDPAEVCOTGGLIYTFEVAORALCPLCGSTGLDPLIFCACCCEPERHQ	1232		
QY	616	FCLTPAGEFFKMGSVHGS-----	KVLLISTLRLTLTOLENNIPSSFLPHPNASHRANMI--	669	
Db	1293	YCVQDEVYLK-----	HGSPFDITLMSILETTYN--ASNGPSSSL--	NQUTQRLMLCP	1343
QY	670	-----	KAYOMCSK-----	PREPALALIECAVAPVYM	697
Db	1344	CTVQYTCNMSSGSKVCKOKCMYHSTCLGTSKRLGLADRPILCYVNLCKJCSCTTKV--	1401		
QY	698	LPIKREFLG-----	-----	HRRL--	710
Db	1402	-----SKFVGLNLPKTCGCFKLTKKGNFCPICQRCYDNDNFDLKMMECGDGOMVHSGEGL	1457		
QY	711	-----HRMTSIEREKEKVKYKKERKQOEETMOQATVWYKTFPVKHQVWKQ--	GEEYRYTG	765	
Db	1458	SDQOYNILSTLPEBIEITCKKCARNNSS-----	KIKAEEMQAMBEFKASL	1505	
QY	766	YGMWSMTKTHVRFVPKLPQNTNVNRYKSLESTKNNMDENDESDDRKCSRPKIKITIE	825		
Db	1506	YSVLKLKSKROACALIKLSPKRNV-----	RCGCGASSNOGKLQ	1544	
QY	826	P-----DSEKDEVGSPAAMGADONEMDISKITEKKDOVYKELLDSDSPCKEKEPMPVD	880		
Db	1545	PKALQFSSGDNGLGSD--	GESONSDDYEFKDDQOO--QOORANNNKP--	RYAPPL--	1595
QY	881	DDMTKTESHVANCQES--SOVDVAVNSEGFLRTSYKKTKSSKIDGLLEBRIKQFTLEEKOR	939		
Db	1596	PCGCOQHISHISQSFILYDI-----	KQKLAGNSYVSLEEFNIDMSQV	1636	
QY	940	LEKIKLEGGIKIGIKTSTNSKNLSESPVYTRAKEGCOSDSMRQOESPANNDQ-----	993		
Db	1637	IQO-----	SNCKDELDIAYK-----	ELLSQDFPMQOETKACTDA	1670
QY	994	-PEDLOGCS-----	QSDSSVLRMSDSBSHTTNKILYKPDRAVLDDVSTIRSEPTKCPKONS	1045	

Db	1671	LEEDMFESC	G	N	V	E	D	L	D	A	G	V	A	S	A	Y	N	H	E	S	T	S	A	E	N	S	G	V	L	D	-----	I	P	L	---	1717																				
Qy	1046	I	E	N	D	I	E	K	V	S	D	L	A	R	G	O	E	P	P	K	S	T	K	C	N	D	E	F	I	D	S	K	L	A	S	A	D	I	G	L	I	C	K	N	K	-----	1095									
Db	1718	-----	E	V	D	F	G	S	C	G	I	K	-----	M	R	L	D	T	R	M	C	L	F	C	R	K	S	G	E	L	S	E	E	A	R	L	L	G	H	D	C	W	H	T	N	C	A	M	W	1770						
Qy	1096	K	P	L	O	E	S	D	I	V	S	S	K	A	L	H	S	V	P	K	S	-----	T	N	R	O	A	T	-----	P	L	S	A	M	1134																					
Db	1771	S	A	E	V	E	E	I	D	-----	G	S	L	O	N	V	A	S	A	A	R	K	M	I	C	V	G	C	R	G	A	T	C	V	N	R	S	C	G	E	H	Y	H	P	C	A	R	S	1826							
Qy	1135	D	-----	F	E	A	K	L	C	O	S	E	S	-----	N	S	T	L	E	N	S	S	D	P	V	S	I	O	D	S	E	E	M	I	V	O	N	S	E	S	I	S	O	F	T	R	E	D	V	1187						
Db	1827	D	C	A	R	L	T	O	K	S	M	C	R	A	H	A	N	G	A	L	A	N	S	P	S	V	T	E	S	N	E	V	-----	S	R	P	V	E	L	D	R	K	-----	1877												
Qy	1188	E	V	E	L	P	K	C	E	L	V	S	G	-----	E	S	T	G	N	C	E	D	R	L	P	V	K	T	E	A	-----	N	G	K	R	P	S	O	O	K	L	E	E	R	1231											
Db	1878	K	L	I	P	A	R	V	O	F	H	I	G	S	L	E	V	R	O	L	G	A	I	V	P	R	F	S	O	S	E	A	V	P	I	N	F	L	C	S	R	L	Y	M	S	K	E	P	-----	W	K	I	V	E	Y	1934
Qy	1232	P	V	-----	N	K	S	O	D	K	L	-----	K	N	T	I	D	K	A	N	N	E	R	S	E	K	G	-----	R	I	S	T	O	I	N	K	D	1272																		
Db	1935	T	V	R	T	I	O	N	S	S	T	I	L	A	D	V	G	R	N	Y	V	D	H	T	I	N	P	S	K	E	V	O	L	G	A	O	I	A	R	M	T	S	L	A	R	E	F	L	E	N	G	T	D	1994		
Qy	1273	N	K	P	I	X	L	G	E	C	L	E	I	S	E	R	V	A	S	G	N	E	P	K	V	N	I	K	I	P	-----	E	N	D	I	K	-----	S	L	T	Y	K	E	S	A	I	R	P	1327							
Db	1995	W	S	G	E	F	P	N	S	C	V	P	-----	D	O	N	T	E	E	E	P	O	O	A	D	L	I	P	P	L	T	S	P	L	F	L	G	L	S	H	G	L	L	M	L	2047										
Qy	1338	N	G	D	V	I	M	E	D	F	E	R	N	S	E	T	K	S	H	L	S	S	D	A	E	G	N	Y	R	D	S	E	T	L	P	-----	S	T	R	E	S	D	T	1374												
Db	2048	L	G	V	V	-----	R	L	K	O	G	E	L	K	D	A	I	F	-----	E	D	L	P	H	E	L	D	I	G	S	M	D	I	F	L	Y	D	K	T	O	L	P	A	I	S	O	S	K	O	T	2101					
Qy	1375	O	T	T	P	S	A	S	C	E	P	S	V	N	O	V	E	M	E	T	E	V	K	A	K	Y	S	S	P	T	S	E	E	N	S	N	D	T	I	D	E	N	G	L	P	T	-----	N	K	N	1433					
Db	2102	Q	A	M																																																				

Db	2590	FVTDGSGGQGLQYISIPAGEKKPQRPQPLATPFLTTAPGAGATVLTQDASGNLVLT	2649
QY	1907	TSNS--QVITPERQIRPEMTYIRPLQOSTL-----GKAIITRVYMQPQAPQVW----	1954
Db	2650	PSNSGLMLTAOSLQAPQVIGTLQIPQITQLGGADGN-----QPSNDQPLITG	2700
QY	1955	-----TQIIRGQPV-----SFAVSAPMTVSSST--PG---OKSLTSA	1985
Db	2701	GTCGGSSGLEFATTSPOVYLATQPMYTGLETTIVQNTVMSQQFVSTAMPGLMSQASFS	2760
QY	1986	TSTSNIOSSASQPERPPOQGVKLTMLQTLQNGHGNGGLTV-----VIOGQGV--TTGQ	2039
Db	2761	TTTQVFGASKPE-----IVDLRAGVYVLNNTGDASSAGTFLNANASVLQOQTDODDTTQ	2814
QY	2040	LQIPQGVTVLPGFGQQLQMAAMPNGTVQREFTPLATTATTAATTTTTSVTAAGTGEQ	2099
Db	2815	I-----LQANFQ--FGSVPTSSGASTSMQDTPSVMTA-----	2846
QY	2100	RQSKLSPQMOVHQDKTLRPAQS--SSVGRPAKQVQ-----TAQSAARPQOTQPSAQ	2151
Db	2847	---KIPPVLTQIKRNNAAQAKAGISGVGVKVPQPVQVNVKVLPSVITQSQVQVKNSLKQ	2903
QY	2152	PEVQOTPEVQOTQTVSSSHVSEA-----OPTHAQSSKPOV-----	2186
Db	2904	SOVKGAASAGTGTCGA--PSTIASKPLQKTKNNIRIHKLEKPKVYMKPTKYNQONHSL	2962
QY	2187	-AASQSPQSNVQSGSPVR--QSP-----SQTRIRPST-----PSLSQSGQS	2226
Db	2963	LQOQOQOQOQPOLQOQIPAVNVVQVPKVTISQQRIPAQTOQOQLQQAOMITHIPQOQOPLQOQ	3022
QY	2227	QVQTTTSPPI---PIQHTSLQIPSGQVQSQVQSSQTQLSSQTLQNVQSVSPSR--	2281
Db	3023	QVQVQSPMPTITTLAEAPVQSGFQVMEQALEQOELANRVQHTSTSSSSSSSSCSLPTNVV	3082
QY	2282	-----POLQIQQPP-----QVIAVPOLQOQVQ--VLSQIQS	2311
Db	3083	NPMQQAASPTTSSSTTRPTKRVLRPMQQRQEPARLSMECSVSSPTPRKPVQDPPIHQMTS	3142
QY	2312	QVVAQIQAQSGVSPQIQIKQLPFIQIQSSAVQTHQIQNVTVQAQSVQEDLQRYQDLRQD	2371
Db	3143	ASVSKCYAQAKSTLSPV--YEAEKLKVSYLE--IVPQVTMDA--ILEQDPVQSIYTE	3195
QY	2372	QQKKKQQQIEIKRHTLQASQNSQETIQQVYMKHNAVIEHLKQKSM--TPAREEN----	2426
Db	3196	GLYKNSPESKTEQQLLQOQREQLNQDVLNNGYLLDHTQVPEMDTVREEDLEE	3255
QY	2427	-----QRMIVCN-----QVMKYVLIDIKDE-----	2447
Db	3256	EDEDDEDSFLMKRTSACNDHEKMSSEEPAYKDKISKLLDNLTDQDASLATAATTMEVDAS	3315
QY	2448	-----KQAAKRRKRESVEQKRSKQNAKTLASLFLKHKREQLRAETLKRA	2492
Db	3316	AGYQWQWEDVLATTAAGASVATEEFEGALETAANVAEAATVAINMDAH-----	3362
QY	2493	LLD--KQIQIEVQSEELKDKLIKKEKQMLQATVAAPACRPVTVLPAAPAPPSPPP	2551
Db	3363	VLDLQQLQNGVELELTRR--KEQORTVQSEQEOSKAALVP--TAAABEPPOPIDQPKKM	3417
QY	2552	PG 2553	
Db	3418	TG 3419	

RESULT 7

T34434

hypothetical protein K06A9.1a - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000

C:accession: T34434

R:Geisel, C.; Gattung, S.

submitted to the EMBL data library, December 1996

A:Description: The sequence of *C. elegans* cosmid K06A9.

A:Reference number: Z21525

A:Accession: T34434
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2232 <GEI>
 A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
 A:Experimental source: strain Bristol N2; clone K06A9
 C:Genetics:
 A:Gene: CESP:K06A9.1a
 A:Map position: X
 A:introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 2.9%; Score 419.5; DB 2; Length 2232;
 Best Local Similarity 21.8%; Pred. No. 5.4e-08;
 Matches 339; Conservative 172; Mismatches 633; Indels 413; Gaps 63;

QY 945 LEGKIGKGTSTNKKNSKESPVTTAKKGCQSDSKROPO---SPNANDOPEDILQGC 1001
 DB 731 LESSTSG-ATTSSGSACTMTSP-----SQSSVSGSGSTSPASATSGEMTSQGS 781
 QY 1002 SQSDSVLRMSDPSHTNKLKPKDRVLVDVSRSPETKCPKONSIEKYSDLASRG 1061
 DB 782 TQFGSSSVTSAAITFTSQ-----QSVSTNSPGSTVTRPST-----VSGSTSSG 825
 QY 1062 QEPKSKTKGNDFIDSKLASADIGTLCKKKPLIQESDPIV-----SSSKALHS 1116
 DB 826 STVTVSGSTES--TSGSSVASSPAPS-7SQNPNPSTSSGSSMTTQSPYPSOSTSPVES 881
 QY 1117 SVPKSTDRDAPVPLSRAMEEGKLGCDSESNSTLENS--SDTVS10DSSEEDMIVQNSN 1173
 DB 882 STTSPSPSGTTLTST-----PSPQSTTIGTQSTSPGISTSEE-MTSGST 931
 QY 1174 ESISEQRTTRQDVEYLEPLKCELVSESGNCDRLPVKGTANGKKPKQOKLEERP 1233
 DB 932 QTPSTGSTVYQPSVSD-----STSSGS--TVVSGSTES--SSPSPSTSQNTNPS 979
 QY 1234 NKCDQJLKLKMTTKNNNESEKKGORTSTFOINGKDKMKPKYLLGCELCIESESRY 1293
 DB 980 TSSSSSMSTQTPQSSQSTSPVESSTSCATSS-----GSP-----GTTLTSTSPSPSP 1027
 QY 1294 S-----GNVEPKVNNINKIPE--NDIKSLVYKESAIRPENGDIYMEDFENRNSSET 1344
 DB 1028 SSTIGSSQSGSTSPVSTISQSTETPGSTGTVPKRPST---VSGSASGSGSTATMGSTEA 1083
 QY 1345 KSHLLSSDAAGNRDSELTSPRK---ESDSTQTTPSAACPENSVNOVEMDEITSE 1401
 DB 1084 SS---TSGSSTSPNPSQSTSPSTSGARSSPGSGTTLTSLSPSQSSTIGSGGSTSP 1140
 QY 1402 VKKVTSPITSEESNLNDPIDENGPIPKNNENVNG-----ESKRKTVT 1447
 DB 1141 VVSTTSGMTSQSGTOIPG---STGSTRVQPTSGSGSTSPSGEITSGSTQTPRSLSLT 1196
 QY 1448 EVTTMTST--VATES--KVIYKVEKDKQTVVSTENCAKSTYTTT-----TV 1493
 DB 1197 SPAISTSTQSGSVSTNSPOSTVTO-----PSTVRGST--SGSTVTTGSGSGSSSA 1249
 QY 1494 TKLST-----PSTGSDIISVKESQKTVVTVTVTDLT-----TTGCT 1532
 DB 1250 TSLSSSVVPSTSQSPNPSSTGSGSTPPNPSQSTSPVSTTGTGEMTSGSTQTPSTIGST 1309
 QY 1533 LVTSMTVKEYSTRDKYKIMKFSRPRKTRSGTALPSYRKFTVYKSTKSIKSYLPPNDLKL 1592
 DB 1310 VTQPTVSGNSSGSTVMI--GSSEASTSGSSSF-----KTSPPSISVPTSS-- 1354
 QY 1593 ARKGGIREVPYFNNAKALIMWIPSPRPF-GITWYKRYQYKSLAGVIMLRMLMAS 1651
 DB 1355 -----PIPSTTFASSTSGSTSDVSVSTSLA----- 1382
 QY 1652 LRMDMAKVPGGGSTRTETSEITETILIKRDVGPYGIKREYCIKICIGVPEI 1711
 DB 1383 -----PLSSSLP-----STVPSSTQSTSSSTSGSSKASSSP-----VPSQ 1417
 QY 1712 PKRPTPORKGLRSSALRPK-----RPTPKQGTGVPVIEETWVAEE 1752

DB 1418 TSTPTNPTGSESTSLSTISGSTQHTMTSKASSGSTSPSTNSQSTVIMG----- 1471
 QY 1753 LELMEIRAPAEVEKEKAQVBOAKKRLQOKPVIATSTTSPSSSTTSTISPAOKWV 1812
 DB 1472 -----SSSTGSGVSTSSASSTQPMOSTSQSSAGSTVASTASPAASSTAPBSTG--TMS 1523
 QY 1813 APISGVTTGTCKMVLTTKYGSP---ATVTFQONKNHOFATWVMOGOSNGSVQVOQKV 1869
 DB 1524 STSSGTVSGSTIESSSTTASASQOTGTVMGSSST-----SGVSTSSASTQPM 1573
 QY 1870 LGIIPSTSTQOQTFSPQRTATVIRPNTSGSGGTATNSQVITGPQIRPQMTVIRPL 1929
 DB 1574 STNSGSSAGSTVASTAGVSTSTV---PSTGTMGSTSSG-----TVGSTIS 1618
 QY 1930 QOSTLCAKAIIRPVMVQPAQVQMTQIIRGQPVSTAVSAPMTVASTPQ-----KSL 1982
 DB 1619 ESSTRTASASQOTGTVTMGSSS-----TSGVSTSSASTQPMOSTSQSSAGSTVAST 1672
 QY 1983 TSATSTNIOSSASQPPROQGVKLTMAQLTQHGNGGLVTVIOGQGT----- 2036
 DB 1673 TGLVSTVPSSTGTVSGTSSGTVSGTISESS--TASASQOTGTVTMGSSSTGVS 1730
 QY 2037 ---TQQLQILPO-----GVTVLPQGOQLMQAAMP-----NGTVORLFTPLAT 2078
 DB 1731 SASGQPMOSTSQSSAGSTVASTASPAASSTAPSTGTMGSTSSGTVG---STMOS 1786
 QY 2079 ATTAATTTTSTTAAGTGEOROKSLSPOMVHODK----- 2114
 DB 1787 STAASTSHSGTVTVLGS-----STSSNOMSTSQSSAGSTVASTAGVSTVPSSTG 1842
 QY 2115 TLPPAASSVGV-----PAKAQPTAO-----PSARPQPTQPOSPAPEVQTOPEVQ 2161
 DB 1843 TMSSTSGTVSGTISSTSTTASASQOTGTVTMGSSSTGVSSTASASTQPMOSTSQSS 1902
 QY 2162 TQTVSSH-----VSEAOPTHAOS-----SKPYAASQPOPSNVQOGSPVRY 2204
 DB 1903 AGSTVASTAGVSTVSTGTMGSTSGTVSGTISSTSTASSTASSTQGTG---STVTI 1958
 QY 2205 QSPSOTRIRPSTPQSOLSPQGOQOVOTTSQPIPIQHTSLQIPSGQPOSPQOVQOST-Q 2263
 DB 1959 GSTSGT--NPSPRSL-----SQT-TITPSPQSTSTQSTSLPSSPSRSTHSVSSSET 2010
 QY 2264 TLLSGQTLN-----QVSVSPSRPQLOIQPOPOVIAVPOLOQOVQVLSQ 2308
 DB 2011 TMSGATSGDKMSFLSTGTGTVSPFSRGSGLATTSAPKPVYCLMPYDQSKRIDQ 2067

RESULT 8
 T34513
 hypothetical protein ZK783.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T34513
 R:Ravello, A.; Vaudin, M.
 submitted to the EMBL Data Library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: 221536
 A:Accession: T34513
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3507 <FAY>
 A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
 A:Experimental source: strain Bristol N2; clone ZK783
 C:Genetics:
 A:Gene: CESP:ZK783.1
 A:Map position: 3
 A:introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1

Query Match 2.9%; Score 413.5; DB 2; Length 3507;
 Best Local Similarity 19.8%; Pred. No. 1.6e-07;
 Matches 263; Conservative 176; Mismatches 469; Indels 417; Gaps 48;


```
QY 1037 ETKCPKNSIENDIEKYSDLASRGOEPTKTKGNDFFIDDSKILASADDTGLTICKNKK 1096
DB 1941 EKKOPREKIE--IDEENSSNSNGOE--KPTTKG----- 1971
QY 1097 PLIOESDPTVSSSKSALH--SSVPEKSTNDRDATPLSRAMEFGKLGDCSESNTLENS 1154
DB 1972 -IVSSSTASISSSTIAEPHVTTISSTITKDMT-----SSKSPENVTWSSS 2018
QY 1155 DTVSIODS---SEEDMIVONSNEISIEQFRTREDEVLEPLKELVSGESTGNCEDRL 1210
DB 2019 PEVSTSSSKSTJASETTVSTPSESSSE-----APLTSSPAT----- 2056
QY 1211 PVKGTGANCKRKSQOKKLEERVYKCSDOIKLNTTDKANNRESEKGCORTSTFQING 1270
DB 2057 -----TTEVTESVSKSTPKESSESEITVKKLSKS----- 2087
QY 1271 KDNKPKIYKGCLEKIESRVSAGVNEPKV--NNINKLIIPENDIKSLVKSAPIRPIN 1328
DB 2088 -----PEVTSSVSKSPSTSTTSOSVSTVETSKSTVLSSEAPVSTSP 2133
QY 1329 GDVIMEDFERNSSSETKSHLLSSDAEGNYRDSLETLPTSKESDSTOTY----- 1377
DB 2134 TEV-----HTSSEKPLSASSTT---GDTNSTPTSSLASVKSSTSAEGTSASVA 2182
QY 1378 -----TPSACSPESNSVNOVEDMEIETSEVKKTTSSPTSEESNLSDNDFIDENGLPI 1430
DB 2183 PVKLSLSLSDVQSPSTKTFDATESSTVQASETSOSGTSVSKSTPESHSVTKLSTISS-- 2238
QY 1431 NKNENVNGESKSKRTVTEVTMTSTVATESKTVIKVEKDKQTVASTENCAKST----V 1486
DB 2239 NPSSSVPTSPKSTPVPPESTIQPTSTTSPSGSLTPMNSNEVLTTSEHVLSSSLSPV 2298
QY 1487 TTTTITVTKLS-----TPSTGSVDIISKESKSTVTTTVDLTTTGTLVTSMTVS 1540
DB 2299 SOSSTTPNNLSSSSVTEPKTSSSEVSLNSEPSTTEAPTLSPDLISTTTNNLSOSSTV- 2357
QY 1541 KEVSTRDKVKKMKFRPKTRSGTALPSTRKTVYTKSKSIVLNPDLIKKLARKGCIRE 1600
DB 2358 ---STEDRSEISENSEKPTSAPELVTSVTHVASSSPD---VPESESPDILTGSSTE 2410
QY 1601 -VPYFNYNKPPALDIMPV-----SPRPFGLIMWRRLQTVKSLAGVSLMLRLMKSRLW 1654
DB 2411 NIP--EASKKQIISTPTPTDTTTASEEPKSTMSMSPDLSTTSVLSSES----- 2456
QY 1655 DDMAAKVPPEGSTRTETSETE---ITTTETIKRDRVPGYIGIRFEYCIKRIICPIGVE 1710
DB 2457 -----STPPESSKSPVSSSTEGISVITSEFSK-----VEE 2488
QY 1711 TPKETPTPQKGLRSALRPKRPETPKQTPVYIETWAVEELE-----LMEI 1758
DB 2489 S-----TISVLEEDLTKT--TPSPILLEETTTASETSEPLTEDSLTVSVRIHEL 2535
QY 1759 RAFAEVEKEKAQAVEQOKKRLBQOKPVIATSTSPSTSTSTSPISPOKWWAARIS 1818
DB 2536 TTSSSEVPEKE--SESTTSSSEKSPSOEPAGILSTVVPVTSVSLTJASE-----IDA 2587
QY 1819 VTTGTAKVLLTKVGPBPAVTFPOONKNFHOTFAWKGQOSNSGVVQVQOKVLGIIIPSTG 1878
DB 2588 ITSNNPPE-----KQGR-----PITTSKPSLVAST-- 2612
QY 1879 TSOQFTTSQOPTATVTPNNTSGSGCTTSSNOVITGPOIRPGMYVIRPPLQOSTLGKAI 1938
DB 2613 TSPSTVTSSEPESESTR-----TTVSTTVST-----TTPTEETTTSESL 2651
QY 1939 IPTPVWVGAGQAVWTOIIRG--QPVSTAVSAPNTVSTPGOKSLTASTSTNSIOSSAS 1996
DB 2652 IIT---AAPSKESTESTSESEAPTPPAKTSKTPSNVSS---TSKSTENETSTIS 2701
QY 1997 QPPRPOQGVKLTMALQTLQTOGHGNOGLTVVIOGOGGTTGOLQIPQGVVLPPGQO 2056
DB 2702 Q-----SCSLE----- 2707
QY 2057 LMOAAMPNGTVOIRFLFTPLATTATTATTTTIV-STTAACTGEOBOSKLS---POMQVHO 2112
```

```
DB 2708 --SSTMSTSSSEETNAPVTVSSSEASTTLEENSTSSPTSSSEASVKLSLPESTISE 2765
QY 2113 DTLPPAOSVSGPAAKAPQTAQAPARPOQIOPQOPAPQEVOTQTOPTVSSHVS 2172
DB 2766 AVTV-----SSNAPAI---TMSSEHREIYSTVSSSPSEPEIPLSTVPPNVYATSIIS 2817
QY 2173 EAQPTHAQSKKPQVAQAQSP--QSNVQGSPPVRYQSPQRIIRPSTPSQL-----SPGQOS 2226
DB 2818 E-----EPILSVTSSSTPRV-----RLITGTPDDLIVSVTVPSHCN 2854
QY 2227 QVQTTSPPIPIOPHTSLDIPQGGQPOQPOVQOSSQTOTLSSQOTLNQVSVSSPRQLOI 2286
DB 2855 RQNTITAVSSPNSNSTPILPSESLTTPQP-PTTTTAKPATTSKGRPPSIQPAEMF 2913
QY 2287 OOPQP 2291
DB 2914 TTPAP 2918

RESULT 9
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence-revision 03-May-1994 #text-change 07-May-1999
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SOCH>
A:Cross-references: EMBL:X65551
C:Superfamily: Kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KIH>

Query Match 2.9%; Score 412; DB 2; Length 2897;
Best Local Similarity 18.5%; Pred. No. 1.4e-07;
Matches 551; Conservative 391; Mismatches 1062; Indels 980; Gaps 137;

QY 291 OCEVCVAHHVPGVTDCAETOKMKPYIRH-----EPIG-----YDR 326
DB 34 ECDIRQLVYVSKQKQKIEHDEALIHNFSSNPTQVNGSVIDEVYRLKHGDIVTIDR 93
QY 327 SRKKYWFLLNRLLIEDTENENEKKIWIYSTKVQVLAELIDCLDKDYWEALCKILEMRE 386
DB 94 SFR-----YENSLQNGRKSTFPPRKIREQEPARVRSFSFSD---PDEKAQ 138
QY 387 EIHRRHMDITEDLTKRKSGNSKSTLAANEEILESTIAKKGIDN-----VKSPEIEND 440
DB 139 DSKVAYSKITE--GKVSQGNPELF---DENLPPNPLKGEAPTAKKSLVMHTPPYLKRI 191
QY 441 KNETENDSKDAEKNRE---EFEDOSL---EKSDDDKT-DDEPEQKSEVGFKSEKSN 492
DB 192 IKEQPPQSKQESGSEIHVEKAQSLVSPAPSPKPTVYASQRRRSCKTAPASSSKQ 251
QY 493 GELSESPGAKGKAGSGSTRITTRLNPDG-----KLSQKSOQVAAAHANKLEKGEKE 546
DB 252 TEVPK-----RQGERVATQCKRVKVSIRSQHDILQICSKRRSGASEAN----- 295
QY 547 VLVVNGGELISLSTK---EVIKMG---NINNYFLQGEQKRYV--YINQYSTN----- 593
DB 296 LIVAKSMADVVKLGAKQOTKVIKHPQPSMKRQRRPATPKRPVGEVHSQFSTGHANSP 355
QY 594 -SFALNKHQHRDHDKRR--HLAHKFLTPAGEFK-----WNGSVHSGKVLITITSL 641
DB 356 CIIITICKATEKVHVARRYVLNNFISQMKDKFEDLSGIAMFKTVPKEDPOL- STC 414
QY 642 RLITQLENNIPSSF-----LHPNVAISHRAMVIKAVQMSCK--PREFALALAILBCA 691
```



```

Db 415 HIAISSENILCKQFOGOTDSGCEPRLPTSESFGGNVEFSAQNAKQPSD-----KCS 466
Qy 692 VKRVVWLPITREFLGHT-----RLHMTSTIERE-----EK 721
Db 467 ASP-----PLKROGIRENGNAVAKTPRNTYKMTSLETSTETEPSTKSTVSTVNSGSTE 522
Qy 722 EKKK--KEKOEEMM-----OATWYKTFPVKHQWAKOGEEYVTVGWS 770
Db 523 RNIOKLPBESSEETETVEICILKRGOKATILQO-----RDEGKEI-----566
Qy 771 WISKTHVREPRLPNTNVYRKSLSEGTKNNDENWESDK-----RKCS--RSPKIK 823
Db 567 -----BRPEET--YKENIELKEN--DEKKAKMRSRTOGQCAPMSDLTLK 609
Qy 824 IEPDSE--KDEVG-----SDAKGADONEMDISKTEKKODVVELLSDSDKCKE--874
Db 610 SLDPTELMKDTARGONLLOTQDHAKAPKSEKKITRM-----PCQSL 651
Qy 875 EPMEVDMDKTESHVNCQESSQDVVNVSE-----GFHLRTS-----911
Db 652 QP-----EPITPTHTKOQLKASLGKVGKELLAVGKFTTSGETHHTHREPAQDGKSTR 707
Qy 912 -YKKTK-----SSKLDGLLERRIKOFTLEKOLEKILEGIGIKGTSTNKNLSE 965
Db 708 TFKESPRKOILDPARAVTGM--KKMPTRPKBEAQSLEDL--AGFKELFQTPGESSEMTD 762
Qy 966 SPVITAKBECCOSDMROESPANANDQPE--DLIOGCSQSDSVLKNMSPSHNTNLYPK 1024
Db 763 EKT--TKIACKSPPEVSPTPTSTKOWPKRSRKADVEEFLAKRLKPLPSAGKAMITPK 819
Qy 1025 DRYLDDVSI-----RSPETKCPKONSIEN-----DIEKVS 1055
Db 820 PAGEDEDIAKFGTPOKLDLAGTLPGSKROLOTPREKQALEDLAGFKELFOTGHTE 879
Qy 1056 DLASRQO-----EPYKSKTYGNDPFLDSDKLASADITGILI--CKNKKPLI 1099
Db 880 ELVAAGKTTKIPCDSPQSDPVDTPSTKOR-----PKRSIRKADVEGELLACRNLMP--931
Qy 1100 QESDSTIVSSKSAIHSVVKSTINDR-----APPLSRAMPFGKIGCDESSSTLENSS 1154
Db 932 -----SAGKAMHTPKSVGBEKDIIIFVGTPOVKL-----DLLENLT 968
Qy 1155 DTVASIODSEEDMIYVONSNESISEOFTRPQDVEVLEPLKCELVSGESTG--NCEDRLLPVK 1213
Db 969 GSKRRPQTPKEAQALEDLGLFKELFQTPGHTEBA-----VAAGTKTKMPESSPPES 1021
Qy 1214 GTFANKKBPQOKKLEPRPVNKCSDQIKLNTTDKKNNRESEBKQORTSTQJINGKDN 1273
Db 1022 ADPTSTRROPKTPLEKRDYOKELSAK-----KLTQTSGETTHTRKVPGEED 1069
Qy 1274 KPKIYJGDEGLKEISESRVAVGNV--EPKVNNIKIIPENDIKSLYVESAIRPINDVI 1332
Db 1070 KSJNAFETKQKLDPAASYTSGRRHPKTK--EKAQPLEDLACM--KELFOTP-----V 1119
Qy 1333 MEDFENERSSETSKSHLSSSDAGNVRDLSLETLPSTKESDSTQTTTPSASCPSNSVNOY 1392
Db 1120 CTOKPTHEKTKIACKSQD-----PYDPTPSKPOGSKSLRK 1158
Qy 1393 EDMEIETSEVAKVTSS-----PITSEESNLMDPFLDENGPLINK--NENVNGES 1440
Db 1159 VDVEEEFALKRTPSAGKAMHTPKPAVSGEKNIYA--FM---GTPQOKLDLLENLGLSK 1213
Qy 1441 KRRTVITEVTMTSTVA-----TESKTVIKVEGDKQTVVSTENAKSTVTT 1488
Db 1214 RLOTPREKAQALEDLAGEKELFOTRGHTEESMT--NDKTARVACKSSOPDLDKNPASS 1270
Qy 1489 TTTTVTLTSPSTGSGVDIIVKESQSTVTTVTVDSLTTLTGGLVTSMTVSKVEYSTRDK 1548
Db 1271 KRRLKTSLG--KYGVKEELLAVGKLTQVTSETHHTHTEPLGDKSKMAFESRKQIILDSA 1328
Qy 1549 VKLMKESRPKTRSGTA-----LPSYRK--FVTKSTKKSIFVLPPNDL--1589

```

```

Db 1329 ASLTGSKROLRTPKCKSEVPEDLAGFIELFQTPSHTKESMTNEKTKYVSARASQDPLVDT 1388
Qy 1590 -----KKIARKGIRVFPVFNNAKPALDIMPSPRPFGITWIRLOTVSLAVY 1641
Db 1389 PTSSKQPKRSLRKADTEE-----ELAFRKOTPSAG--KAMHTPKPAVE 1432
Qy 1642 SLMLR--LWASLRDMDMAKVPBGSGSTRTESEETITTEIKRRDVPYGR-----1694
Db 1433 EKDNIFLGTPOVKLDQPNLPR--GSNRRLOTREKAQALEELT-----GFRELFOPT 1483
Qy 1695 -----FEYCIRKILIC--PIGVP--ETPKETTPPOKGLRSS-----ALPKRP--1733
Db 1484 CTDNPRADKTKTKIICKSPQSDPADTPNTKORPKRSILKADVEEFLAFKRLPSACK 1543
Qy 1734 --ETPKOTGPVITETWVAEELELWEIRAF--ARVE-----KEKAQ 1772
Db 1544 AMHTPKAA-----VGEK--DINTFVGTPVKLDLGNLPGSKRRPQTPREKAKA 1591
Qy 1773 VEOAKKRLQOQKPTVIASTTSPSTSTISPAQKVAVAPISGVTGTGKMLTTKYG 1832
Db 1592 LEDLAGFKELFQTPGHTEESMTDKITEVCSKSPQDPVKTP-----TSKORLKSIG 1645
Qy 1833 SPATVTFQONKNPHQTFATWAKOGNSGVVOYQKVL--GIIPSTGTSOQTFPSFOR 1890
Db 1646 K-----VGKKEEVLVGLTQTSKTTQT-----1669
Qy 1891 TATVTIRPNTSGSGT-----TSNSQVI-----TGPOIRPGMTVTRTP--LOQST 1933
Db 1670 -----HRETAGDKSIRKAFKESAKOMLDPANGTGMEMP--RTKKEAQSLEDLA 1718
Qy 1934 LGRATIRTPVWOPGAPQOVMTQIIRGQPVSTAVSAPNTVSTSGOKSLTSATSNIOS 1993
Db 1719 GFKELFQTPDHTEESTDDTKTKI-----ACKSPPESSMDTPTST-----1758
Qy 1994 SASOPRPOQGVKTL--MAQLTQLTQHGNGNGLVIVOGQQTGQQLILP--OGV 2047
Db 1759 --RRRKTPDKGRDIVEEISALKOLTO-----TTHDKPVGDEDKGI 1798
Qy 2048 TVLPGPGQOIM-----QAAMPNGTVQ-----RELF--TPLATTAFTTASTTTT 2087
Db 1799 NVRETRAKQKLDPAASYTSGSKRPRTPKGAQPLEDLAGKELFQTPCOTDKRTTHEKTT 1858
Qy 2088 TVSTTA-----AGTGE--QROSKLS--POMOVHODKTLPPAOSSSVGPAAKQOTA-----2134
Db 1859 KIACRSPQDPDPVGTPTIFKQSKRSILKADVEEESLALRKRPVSKADMTPKPAGDCK 1918
Qy 2135 -----QPSARPOQTOPQSPAPQVQPEVQOTQTVVSS-----HVPSEAO 2175
Db 1919 DMKAFMGTPVOKLDLPGNLPGSKRMFOTPRE-----KAQALEDLAGEKELFOTPGTK 1971
Qy 2176 PTHAOSKRPVAAQSQPOSNVQOSQSPVROSPQTRIRPS-----TP 2217
Db 1972 PT--TDEKTYKIAKLS--PQ-----PDVDTASTKORPKRNLKADVEEFLALKKRP 2022
Qy 2218 S-----QLSPGOQSOVQTTTQSPPIQPHTSI--QIP--SOGQPOQOP-----2256
Db 2023 SAGKAMDTPKPAVSDENKINTFEVETPOVKLDLGNLPGSKROPQTPREKAEALEDLVGK 2082
Qy 2257 ---QVQSSQTQLSSGOTLQNVSVSSPSRPOLOQOQPOVYIAP---QLOQOVYLSOI 2309
Db 2083 ELFQTPGHTEESMTDKITEVCSKSPQESFKTSRSKORLKLPLVVKMKEEPLAVSKL 2142
Qy 2310 --OSOVAQIOAQOQSGVPOQIKL--OLPQIOQSSAVQTHQIONVT--VOAASVCEQOLR 2364
Db 2143 TRTSGETTQHTHPTGDSKSIKAFKESPKOILDPASAVTSGSKROLTRREKARALEDVLD 2202
Qy 2365 VQOL-----RDQOQK-----KQOQOIEIKR-----EH 2386
Db 2203 FKELFAPRGHTEESMTIDKTKIIPKCSPPPELDTATSTKRCQTPRREVEKELSAVER 2262
Qy 2387 TLOASNOSEIIOKOVVKNHNAVTEHLKO--KSMYTAERENORMTVYCNQWAKYIILDKID 2444
Db 2263 LQTSQOSTHTHREKAPASGEG--IKVLKORAKKPNVVEEPPSR-----2305

```


Qy	2445	KEEKQAKKKRRREEVEQKRSQNTKTSALLFKKEQDRA-----EILKKR	2491
Db	2306	-----RPRAPKEKAQPLEDIAGFTLESETSGHIOEELTAGKATKIPCESEPPLEVDTT	2358
Qy	2492	ALLDKDIDIOEVELKRDLTIKKEDIMQLO-----ATA	2526
Db	2359	ASTKHNHLTRVOK-----VOVKEEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESKQ	2413
Qy	2527	VAAQCPPTVPVLPAPRAPPSP-----PPRQVOTGLSTPTLPVSAQKRRKEEK	2578
Db	2414	TPAPAASTVTSGRRRRAPRAREQAIEDIAGFKDRPAGHTESMTDDKTKIPCKSPLE	2473
Qy	2579	DSSSKSKKKMISTSKETTKDKTYLCICKTPYDSEKEYIGCDRCOMNYHGRCVILQSE	2638
Db	2474	DTATSSKRPRPTRAQKVEKEE--LLAAGK-----LTQTS	2506
Qy	2639	AELIDVYVCPOCQSTEDAMVLTPLTERKIDBGLKRVLSLQAHKMAWPLEFVDPN-DAP	2697
Db	2507	GE-----TTHTDKEPVEGEGKTKA-----FKQAPAKRNVDAE	2537
Qy	2698	DYGVYIKPEM-----DLATHEEVQR-RYIEKLTPEVAD	2730
Db	2538	DVYTSGRQRPAPRKEKAQPLEDIASQEOISQTPRGHTEELANGAD	2581

RESULT 10
T21435
hypothetical protein F26H11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <WIL>
A:Cross-references: EMBL:Z81515; PIDN:CA854234.1; GSPDB:GNO00020; CESP:F26H11.3c
C:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
;313-368/Domain: bromodomain homology <BRO>

	Query Match	2.8%	Score 404.5;	DB 2;	Length 457;
	Best Local Similarity	24.6%;	Pred. No. 2.6e+08;		
	Matches 105;	Conservative 82;	Mismatches 159;	Indels 81;	Gaps 10;
OY	2360 EOLOR-VQOGLRDOOKKKKOOOEI-----KREHTLQASNGSEIIQOKVYKHNHVI---	2409			
Db	21 ELRLKEIKRRTRMENEEQOGCLTTPRKRAAPRHAKPAEYKKIVINPADITTLGSD	80			
OY	2410 --EHLTKKSMTPAREENNOBIAVCNQMYKYLIDKIDKEEQQAARRRRSEVEQRSKQ	2467			
Db	81 TYDYVKEEKPPESTATNVSRRRRTSANKLSKSBDDR-DKPEQOSTPKKERKTSPASH	139			
OY	2468 NATKLALLFKHEQLRAELIKKRALLDLDQIEVOEELKRDLIKKEKDLMOQLAQTAV	2527			
Db	140 VAFPTGSATPHDILNLSTEHCTCKIFPAS-KLYIQCEL---CARWYHGDCVGVAEQITL	195			
OY	2528 AARCPVPVPLRAPAAPPPPPPVGVOHTGLSTPTLPVASQKKRREEKDSSSKSKK	2587			
Db	196 -----GLSHW-----SCEEICEIOER-----	211			
OY	2588 KMISTSRETKKDTFLYCICKTPYESCFYGCDRCQMWMYGRGCYLIQSEAEILDEYVC	2647			
Db	212 -----VDQDPALTYCQKPDDIDTFYVGDCSQCGMFHFECGGTTTRAEEQCADYNIC	262			
OY	2648 PQC-----OSTEDAMTVLPLEKDEYEGKLRVLRSLOAHKAMPLEFEPDPDA	2696			

[illegible]

RESULT 11
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R:Schluener, C.; Duchow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
U. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.

A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1*3256 <SCH>
A:Cross-references: EMBL:X65550; NID:9415818; PIDD:CAA46519.1; PID:9415819
C:Superfamily: Kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: kinase interaction domain homology <KH>

Query Match	2.8%	Score 404.5;	DB 2;	Length 3256;
Best Local Similarity	18.6%;	Pred. No. 3.2e-07;		
Matches 546;	Conservative 378;	Mismatches 1049;	Indels 955;	Gaps 136;
QY	339	IIIEPTDENNEKKIMVYSTKVQVLAELIDCLRK----	DYWAEMELCKIIIEEMBEIHRHNDI	394
Db	432	VIPTETELIHNEFFLITLWLTQYVERKIQKDSLSPREKLGTTAGOMCSGLPGLSS-----	VDI	486
QY	395	TE-DLTKKARG-----SNKSFLAANEELDE-----	SIRAKKGDIDN-----	VKSPEE 436
Db	487	NNFGDSINSESGIPLKRRRVSEFGHLAPRELDENLPRPTPLKRGAPATPKRKSIVMHPRV		546
QY	437	TEKDKNETENDSKDAEKNE--EFEDOSL-----	EKSDDKTP--DDDEQCKSEYGDKS	488
Db	547	IKRIIKKEPOPBGKQESSEIHVEYKAQSLVISPAPAPRPTPAASDORRRSCCTAPASS		606
QY	489	EKSNGELSESPQAGSGASTRIIFRLRNPPS-----	KLQOLSQOYAAAHANKLFK	542
Db	607	SKSQIEVRK-----RGERVATCLOKRVYSISMSQDIIQIMCSKRSGASEAN----		654
QY	543	EKEEVLVYNQSGEISRLSTKR--EVIMKG--	NINNYFKLGQEGKYRV--YHNOYSTN-	593
Db	655	-----LIYAKSMADYVKLGAKQIQOTQVKVIKHGPOBSMNKRQRRPAPRPPKRVGEVHSQFSTGH		710
QY	594	-----SFLANKHQHEDDDKRR--HLAKFCCLTPAGEK-----	WNSYVSGSKYLT	637
Db	711	ANSPTIITIGAAHTEKIVHVPAPRYRVLNNEISNCKMDEKEDLSGIAEMFKTPVVEQDQPT		770
QY	638	ISTRLTLITOLENNIPSSF-----	LHPNMASHRANWIKAYOMCSK--PREPALAI	687
Db	771	STCHIAISNSBNLKGKPOGTDSEEPFLPTSESEFGGNVFFPSQONAKOPSD-----		822
QY	668	LECAVKPYVMPLPIWBEPLGHT-----	RLHMTSIERE-----	719
Db	823	-KCSASP-----PLRQOCIRENGNNAKTPRNTYKMTSLSTKTSQDETPEPSKTVSTVNSGR		877
QY	720	--ERKAYVK--KEKQOEELTM-----	QOATWVYKTPPVKHQVWAKQKQGEYRYTGY	766
Db	878	STEFNNIKLPEVSKSEETNELIVECLIKRQOKATLLQO-----	RRGEKKEKTI--	925
QY	767	GGWMSISKTHYVREVPKIPGNTNVYRKSLGCTKNMMDENDESDK-----	RKCS--RSP	819

Db 926 -----ERPET-----KENTELKEN--DEKMAKMSRTWGOCAMSDL 964
 QY 820 KIKIIPDSE--KDEYKG-----SDAAGADONEMDISKITEKKDQVKELLSDSDKP 871
 Db 965 TDKSLPDLFELMDQFARGOULLQTOHAKAPKSEKGIITKM-----P 1006
 QY 872 CRK-EPREVDMDKTESHVWCSSSOVDVNVSE-----GFHLRTS-----911
 Db 1007 COSLOP-----EPINTPTHTKQOOLKASLGKVGKEELLAVGKFTRTSGETHTHHREPADG 1062
 QY 912 -----YKKKTR-----SSKIDGLERRIKOFTLEEKORLEKIKIEGKIGKIGTSTNSK 961
 Db 1063 KSIIRTESKQOLLDPAARTGM--KKMPTPKHEAQSLEDL---AGFKLFQTPGSPSE 1117
 QY 962 NLSBPVITAKAGCOSDSMRQOSPANNDOP--DLIOGCSOSDSVLRMSPSHTTK 1020
 Db 1118 SMTDEKT---TKIACKSPPEVSVDFTPTSTQWPKRSLRKADVEEFELAJLKLTPSAGAM 1174
 QY 1021 LYPKDRVLDDVSI-----RSPETCPKONSLEN-----DIE 1051
 Db 1175 LTRPAGGDEKDIKAFMGTPVOKLIDLATLPSSKROLQTPKEKAQLEDLAGFKELFQTP 1234
 QY 1052 EKVSIDLASRGQ-----EPTKSKTKGNDEFIDSKLASADDIGTLI--CKNK 1095
 Db 1235 GHTEBELVAAGKTKIKPCDSPQSDPVDPTSTQOR-----PKRSIRKADVEGELLACRNL 1288
 QY 1096 KPLIOESDITVSSSKALSSVPKSTNDRD-----ATPLSRAMDPEGKIGCSESENSTL 1150
 Db 1289 MP-----SAGKAMHTPKPSVGEKDIIEVGTPOKL-----DLT 1323
 QY 1151 ENSSDTVSIOSSSEDMIVONSNESISEOPRTREQDVELEPLKELCEIVSESTG--NCEDR 1209
 Db 1324 ENLTGSKRRQOTPREKAQLEDLGFKELFQTPGHTPEA-----VAAKTKTKMCESS 1376
 QY 1210 LPVKGTEANGKKSQOKLEERPVNKCSDQIKLKNTTDKNNRESEKKGQRTSTFOIN 1269
 Db 1377 PPEADPTSTRTROPKTPLEKRDQOKELSAK-----KLQOTGGETTHHDKVP 1424
 QY 1270 GKOKKPKIYLKGECLKESISRVASGVN--BPKVNNIKIIPENDIKSLVJESAIREFIN 1328
 Db 1425 GGEDEKSNABRETKAKLDPAASVYTGSKRHPKTK--EKAQPLEDLAGW--KELFQTP-- 1477
 QY 1329 GDVIMEDPFNERNSSETSHLLSSSDAGNYSLETLPTSTRESDSOTQTTSPASCPESNS 1388
 Db 1478 ---VCTKPTTHEKTKIACRSQD-----PVDPTPSKPSKR 1513
 QY 1389 VNOVEDMEIETSEVKVYTS-----PITSSEBSNLNDPIDENGFLINK---NENY 1436
 Db 1514 SLRKVDVEEEFFALRKRTSPSAGKAMHTPKPAVSGEKNIVA--FK---GTPVOKLIDLENTL 1568
 QY 1437 NGEKRTKVTITEVTTMTSTVA-----TESKTVIKVEGDKQOTVAVSTENCACS 1484
 Db 1569 TGSRRRLQTPKEKAQLEDLAGFKELFQTPRGHDEESMT---NDTKTAVACKSSOPDIDKN 1625
 QY 1485 TVTITTTTTVTKLSTPTSGSVYDIISVEOSKTVTTVTLSLTGTLVTSMTVSKEYS 1544
 Db 1526 PASKRRLLKTSLG--KVGVKELLAVGLQTSGETTHHTPEPGDGSKMAKAFESPQOI 1683
 QY 1545 TRDKVAKLMKTSRPKKTSRGA-----LPSYRK--FVTSYTKKSIFVLPEND 1588
 Db 1684 LDSAASLTGSKROLRTPKGSEVEDLAGFLELFQTPSHRKESMTNKTTKVSRASQPD 1743
 QY 1589 L-----KTLARKGIREVPYFNYNNAKPALDIMPSPPTGIMRWYRLQTVKS 1637
 Db 1744 LVDTPTSSKPOPKRSLKADTEF-----EFLARKQOTPSAG---KAMHTPKP 1787
 QY 1638 LAGVSLMLR--LWASLWMDMAAKVPPGGSTRTETSETITTTTETIKRNDVGYGIR-- 1694
 Db 1788 AVGEKENDINFLGTPVOKLQDPGNLP--GSNRRLOTRKREKAQALEELT-----GPREL 1838
 QY 1695 -----FEYCIRKIIC--PIGVP--EPPKETPTPQKGLASS-----ALRKRP 1733

Db 1839 FQTPCTDNPJADEKTTTKILICKSPQSDPADPTNTKQREKRSKLKADVEEELAEFRKLTP 1898
 QY 1734 -----ETPKQGPVILIEFWAVEEELMEIRAF-----AERE-----KE 1768
 Db 1899 SAGKAMHTPKAA-----VEEEK---DINTFGTPEYKEDLDLGNLPGSKRRPQTRKE 1946
 QY 1769 KAOAVEQAKRRLQOKPVTIVATSTTSPSTSTTSPAKVMVAPISGSVTTGKMYLT 1828
 Db 1947 KAKALEDLAGFKELFQTPGHTHEESMTDKITEVSCKSPQDPVKTP-----TSKQRLK 2000
 QY 1829 TVKVSFATVTPQONKNFHOTFATVWKQOSNGSVYVOQVKV--GIPESSTQOQTFIS 1886
 Db 2001 ISLIGK-----GVKKEEVLVPKLTQTSKTTQOT--- 2028
 QY 1887 FQPTRTVTRIPNTSGSGT---TSNSQVI-----TGPOIRPGMTVIRTP---L 1929
 Db 2029 -----HRETADGKSIKAFKSAQOMLDPRANYGEMERW---RIPKREASL 2073
 QY 1930 QOSTLKAKIIRTPVWVOPAQOVMTQIIRGQPVSTAVSAPMTVSTPQKSLTSATSTS 1989
 Db 2074 EDLAGFKELFQTPDHTEESTDDKTKI-----ACKSPPESEMDPTST- 2117
 QY 1990 NIOSSASQPPRQOQGVKLT--MAOLTOITQGHGNOGLTVIYQGGQTTGLOLILP--- 2044
 Db 2118 ---RRRKTPPLGRKDIIVEELSAKLOLTQ-----THTKVPGDE 2153
 QY 2045 -OGVTVLPGRGOOLM-----QAAMPNGTVQ-----RPLF--TPLATTATAS 2083
 Db 2154 DKGINVFRETAKOKLDPAASVYTGSKRQTPRKGAQPLEDLAGIKLFQTPCTOKPPTH 2213
 QY 2084 TTTTIVSTTA---AGTEB--OROSKLS--PQMOVHODKTLTPPAOSSSVGRPAKQOTA- 2134
 Db 2214 EKTTKIACHSPQDPVGTPTIEFPQSKRSRLRKADVEEESIALRKRPVSGKAMDPTPKAG 2273
 QY 2135 -----QPSARPOQOTQPSAPQOEVOQTOPEQOTQTVSS-----HVP 2171
 Db 2274 GDEKDKMAFMGTPVOKLIDLEGNLPGSKRMPQTPKE-----KAQLEDLAGFKELFQTP 2326
 QY 2172 SEAQPTHAOSSKPOVAOSQPOSNOVGOSPVVQSPQSTRIRPS----- 2215
 Db 2327 GMTKPL--TDEKTKTKIACKS--PO-----PDVDTPTASTKORPKRNLRKADVEEELALR 2377
 QY 2216 --TPS-----QLSPQOOSQVOTTTQOPITPQHTSL--QIP--SQGPOQOP----- 2256
 Db 2378 KRTPSAGKAMDPTPKPAVSEKKNINTEVEPVOKLIDLGMLPGSKRQTPPKKEKALEDL 2437
 QY 2257 -----QVQSTQTLSSQOTLNQVSVSPSPQOLOOPQPOVIAVP---OLOQOVQV 2305
 Db 2438 VGEKELFQTPGHTHEESMTDKITEVSCKSPQSPSEFTSBSKQRLKIPLVKDMKEPLA 2497
 QY 2306 LSOI--QSOVVAQIOAQOSGVPQOIKL--QLPLOIOQSSAVQTHQIONVVT--VOAASVOE 2360
 Db 2498 VSKLTRTSGETTQTHTEPTDSKSIKAFKESPKQOILDPASAVTGSRRLOTRKREKARALE 2557
 QY 2361 QLOROVOL-----RDOQOK-----KKOQOUEIKR--- 2384
 Db 2558 DLDVDFKELFAPGHTEESMTIDKNTKIPCKSPPELDTATSTKCPKTRPREVEEELS 2617
 QY 2385 --EHTLOASNOSEIIOKOVYMKHNAVIEHLQ--KKSMTPAEREENORMIVANOYKUYLT 2440
 Db 2618 AVERLTQTSQOHTHKEPERSGDEG--IKVIAKQAKKPNVVEEESRR----- 2664
 QY 2441 DKIDKEQAKKRRKEESVEQRRSKOMATKLSALLFKHKEOLRA-----EL 2487
 Db 2665 -----RPARPEKAQPLEDLAGFELISETSGHTQOESLTAQKATKIPCESPPELV 2713
 QY 2488 LKKRALLDKDLOLEVOELKRLDKIKKENDLMOAQ----- 2523
 Db 2714 VDTTASTKRLKLRVOK-----VOVKEEPSAVKFQTSGETTADDEPPAGEGKITALKE 2768
 QY 2524 -ATAVAPCPVPVTVLPAPPAPPSP-----PPEGVQHTGLSTPTLPVVASQKRRR 2574
 Db 2769 SAKOTPARPASVYGSRRRRPARAPRESAQALEDLAGFPDPAAGHTEESMTDQTKTKIPCKSS 2828

Qy	1338	ERNSETEKSHLLSSDAEBENYRDSLETPSTKESNSDTOTTPPSACGESNSVOWDMEI	1397
Db	1484	EQSDSESTATIRRDPAQ-----KTEEISKODDGEKTTT-----DCKPRKPDSEA	1531
Qy	1398	ETSE--VKYVT--SSPTSE-----EESNSJNFIDENCLPINKNENYCESKRRVITEY	1449
Db	1532	TPKRRVYKKTQKSDSVASDASLADYSKLSDVEEKPKKYLKKT-----EKDSVISE	1586
Qy	1450	TYMTSTVATESKTIVIKVEGDK-----QIVVSTENCAR-STVTTTTTVTKLS7	1498
Db	1587	TSSVDTIKPEES-VEIPTEKAEQMIHNRSTSDSAVESEPKNAHKDDTEKTDIMMRKS	1645
Qy	1499	PSIOGSVDIISVQSKTVTT-----VDSLTGTGT-----LVYS	1536
Db	1646	SA-----IFSDQSSISKTSSEGRRRRRRTGPAKSAOTLALRGDWEIEALLAED	1699
Qy	1537	MTVSEKSTEDK-----VKLMKFRPKRTSGTALPSYRKFTVSTKSFV	1583
Db	1700	DTVMKXNKGADADLNSCHMSHTPERTLIDEVPTDSGEIATATGESHHTLKEVE	1759
Qy	1564	LPNDLKLARKGIRE-----VPPFNNAKPAIDIMPYSPRTGJITMYRLQTVKS	1637
Db	1760	LPVFFVYVLPKRTSGKGEQEVITISVTLNHPIDISKVW-LDQGR-LEINKDSIDTVG-	1816
Qy	1638	LAGVSLMLRLTMAJLRMDMAA-KVPPGGSGTRT-----ETSESEITTT	1680
Db	1817	-CNSVTLR-----RAKEDSGKKVYCDVDCGSHJSTQGRVYLNKMSETKPVITVDKD	1871
Qy	1681	-EIIKRRDGPV-----GIREFYCIKRIICPI-----QVETPK--EIPTP	1718
Db	1872	QFSLILVADSNPBEASFMTVDGKDLDFEGRSRIDVDDGLKLRKGVSKTAGEVEYKL	1931
Qy	1719	QKGLRSSALMKPKPEPPTQTVPIE-----TWAEELTWEITRAFARVEKEK	1769
Db	1932	NEGEVAAKFEVYKNDPSPAPGDVSVYKASCDLHEMTAPEDNGAETVSYV--IEKE	1989
Qy	1770	A-----QAVEQOAKKR-----LEQOKPTVIASTTSP-----TSSSTSTISPAQ--	1808
Db	1990	SGRRKFHVATVNGKKTSTYVDDLEIETPIYIAVANRFGEGFETKRPVOTGSPPOVP	2049
Qy	1809	KYMAVPIGSGVTT-----GTRKVLITRVG	1832
Db	2050	TYEPPTIDVNTJSTCSLSMPKPIEDGSGVYVGYDYVKKENGEWOMKANGEVLFTESEFN	2109
Qy	1833	SPATYTPQONNFQOTATWVKQ-OSNSGVVVOQKVLGIPS-----STGTS	1880
Db	2110	VRA---LSSGKEYEKEIEACMEAGLRBSNVVSKLTVEGLVEPIILDMPVKVLNDKV	2166
Qy	1881	QOTFTSFQPRATVYIRPNISG-----SGGITSNQ-----VITG-----	1915
Db	2167	EVTWKSDEKEFEVVOYKSDGSSIMASVYDGGPSESAATSKCITDGLRBJDPYFVRVAA	2226
Qy	1916	-----POIRGMYVITPQOOSTL-----G	1935
Db	2227	NQHGGESESEPIIPVYLADAPRVLKAIKPVITPKKGEIRLECHAAHGAPEVIMYKDG	2286
Qy	1936	KALIRT---PVMQOPAQOVMTOIRQPVST-AVSAPVYVSTPQOKSLTSATSTSI	1991
Db	2287	KEIIPDENTEIEVNGSMALIIHELAGEVGLYKVLVENINGTETSEAEVGISDVRAHF	2346
Qy	1992	QSSASQPPRPOQ-QVKLT-----MAOLITOL-----	2016
Db	2347	NSSFELTEIEBGHDIETCEVSDEEAVVMYKDGKVLASDRVQFYAMAKRTILRIKGS	2406
Qy	2017	-----TOGHGNGOGLTV-----TGQGT-----TGQ	20399
Db	2407	TJADSGVYKCEITDGRSTEGEVIYVNDQEPHILVGPDAIVKDFGETMVLFCETSKPYAK	2466
Qy	2040	LQIIPQCVTVLPGRQOQLMQAMANGVQKFLPTPLATATTAATTTTIVSTTAAIGTQEO	20999
Db	2467	VYKFKNGEIVWQNKRAIEMDGRATLEIKNDKHDIGATVYASKEKTSAPALVFEV	25266

QY	2100	ROSKLSQ----	MOVHQDTLP	PPAOSSSVGAKKAPQ	TAQPSARPPQQTQ	POSPAPAEV	2154	
		: :	:	:	:	:		
Db	2527	APNLIIFETIR	DGVTA	WAGNEPFA-----	:	VERSGEPIITHL	TNNGNPLK 2572	
		:	:	:	:	:		
QY	2155	QTOPEVOQ	OTTVSSHVSEAP	PHNAOSKROYAAGSOP	OSNVOGSP	PRVOSPSOTIRP	2214	
		:	:	:	:	:		
Db	2573	AIANVTEIDDS	VSVRM--	KDVTLDNSG	IVRYIAES	PLGQCIRKIP	PLKIID-----KP 2622	
		:	:	:	:	:		
QY	2215	STPSQSPGQ	SOQVOTTT	SQPIRPIOPHT	SLQIPSGQ	POPOFOSV	STQTLSSGOTLNOV 2274	
		:	:	:	:	:		
Db	2623	SAPCDL----	QFKVETEDSV	FLSMOP-----	PLETNGAP-----	:	LTGVIERK 2661	
		:	:	:	:	:		
QY	2275	SVSSES	-RPOLQIOPOP-----	QVIANPOLQOQ-----	VOYLSQISOQV	2313		
		:	:	:	:	:		
Db	2662	AVDNRMKRP	CCQVYKPTKLT	VEADELFC	NOYGFRLI	AVNEVESEPC	DTVDVLTLESSEP 2721	
		:	:	:	:	:		
QY	2314	VAQIOAOSG	VPQIKLOLP	PIQIOSSAV	OTHQICWVIV	QAASVOE	-OLOVQOLR--- 2369	
		:	:	:	:	:		
Db	2722	VS--	ESSELPYK	AIILRTP--	QVIYAVD--	EIKVTLRNE	CECPETISLVKVEKKGD 2772	
		:	:	:	:	:		
QY	2370	-----	DOOQKROOQ	IEIKREHTLQ-----	ASNOSETI	OKOVYMKH	NAVIEHLK 2413	
		:	:	:	:	:		
Db	2773	SDWLEIANT	DKNKFRDS	SLFESGEYU	OVATGITH	AVSSPSEET	PNVYKILVPGSEMPASK 2832	
		:	:	:	:	:		
QY	2414	OKKSMT	PREBENOM	IVNOVMYKIL	DKIDREKOA	AKRRKRES	VEBOKRBNQMATLS 2473	
		:	:	:	:	:		
Db	2833	TEKTI	DAKSESEKS--	AEEIV-----	AEKOVDO	SQASESTTE	VEVEKKTK----- 2877	
		:	:	:	:	:		
QY	2474	ALLFKH	QOLRAEIL	KRKALLDK	LDLOI--	EVQBELK	RDKLIKKEKXOJLMO	LAQATAVAAPC 2531
		:	:	:	:	:		
Db	2878	-----	KVKKKVA	NKKEBELLQ	VEKELIKK	GAIVEKU	QODSRSGLQAS-- 2922	
		:	:	:	:	:		
QY	2532	PEVTEVL	PAPAPAPP	SPPPPGV	ONTGLST	PTPLIV	ASQAKRKEE	KSDSSSKKKKMTS 2591
		:	:	:	:	:		
Db	2923	-----	DNESYTT	TSERKS	SEALEK	NSEKSA	EKTS 2953	
		:	:	:	:	:		
QY	2592	-----	TTSKET	KKDT 2601				
		:	:	:	:	:		
Db	2954	ADLEAD	AKRAET	ETKSET	GKET 2973			

RESULT 13
A53577
ascites sialoglycoprotein 1 - rat (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C:Accession: A53577
R:Wu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A:Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, b
A:Reference number: A53577; MUID:94216302; PMID:163496
A:Accession: A53577
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1630 <MUA>
A:Cross-references: GB:U06746
C:Keywords: glycoprotein

Query Match	2.8%;	Score 402;	DB 2;	Length 1630;
Best Local Similarity	19.8%;	Pred. No. 1.6e-07;		
Matches 296;	Conservative 197;	Mismatches 620;	Indels 382;	Gaps 56

```

QY 1205 NCDNRDLYVGSTENOKK-----P50CKKLEBPVNRKCSOQIKLTKT-----TD 124
      21 SCILLIPVMTSTTTSAPKRTSTALPSTNSQWTSNPT--ASSRYMKTNGQASPMVTS 78
QY 1248 KNNENRESEKKGORTSTFOJNGKNNKRTYLKBECLKEISESNVYSGNNEPVNNINKI 130
      79 SSITTLPOSQHGSMKTT-----RNQOTGTTEVTTLTTSAS--SDQVQVETTSQTL 129
Db
QY 1308 IPENDIKSLTVEESALRPINGDIWTEDEFNERNSESEKSHLLSSDAEGNRTDSELTEDS 136
      130 SPQTTTTSNAPRESPP-----STSVILITLTASTEGTSGGTGHMAV 172
Db

```


Query Match 2.88; Score 401.5; DB 1; Length 1367;
Best Local Similarity 21.28; Pred. No. 1,4e-07;
Matches 289; Conservative 166; Mismatches 564; Indels 339; Gaps 56;

Y 2320 OOSGVPQOIKLPLPQIQOSSAVOPTHQIQNVYQAAVSGDQLRQOQKKKKQO 2379
Db 1090 TTS-----TLSFSNGSTQTOTVSMETSSSGKITAPSTSSRRTPVANSDFITTD 1141
Y 2380 IEIKREHTLOASNOSEIIQOKVNMKNNAVIEHLKOKKSMTPEAREENDRMATVQNMVKYI 2439
Db 1142 TSGNAGHTLLGSHSVLTSR-----VASTLGRSTVAHHQSTGRSSTSHQ--SYL 1190
Y 2440 LDKIDKEKQAAKKRRKEESVEOKRSKOMATKLSALFKHKEQLRAETILKRALLDKQ 2499
Db 1191 TESMG-----ASSTSETSLTTEATETSETCILEPTWTH-----CDRL- 1226
Y 2500 IEVGEELKRDLIKIKKKEDLMQAOATAVAACPPTVPALPAPPAPEPPPPPPG 2554
Db 1227 --LSMTSSSGLTTKTDNDRSTALSTSLTLAPPS-TTASRSTVPPAPLPDPQGI 1278

RESULT 14
548478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Yeast (Saccharomyces cerevisiae)
C:Alternate names: extracellular glucanamyase; mucin-like protein MUC1; protein YIR01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text-change 12-Nov-1999
C:Accession: S48478; A26877; S27281; JG6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: EC:247047; EMBL:238061; NID:g603997; PID:g763364; GSPDB:GN00009;
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600; PMID:3106330
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 762-1331 <YA2>
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A:Reference number: S27281; MUID:89031230; PMID:3141213
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA33069.1; PID:g4552
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseud
A:Reference number: JG6123; MUID:9633237; PMID:8710886
A:Accession: JG6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAM>
A:Cross-references: GB:U30626; NID:g1304386; PIDN:AA49609.1; PID:g1304387
C:Genetics:
A:Gene: SCD:MUC1; STA2: MAL5; DEX2; SCD:S0001458
A:Cross-references: MIPS:YIR019C; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:33:37 ; Search time 12.712 Seconds
(without alignments)
3426.253 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 1433

Sequence: 1 MVSEEEEDDDEEDDAETDSE.....KIKGFKASRSHNNKLOSTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100480 segs, 15661496 residues 100480

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1474	10.3	328	US-09-925-297-816	Sequence 816, App
2	982.5	6.9	238	US-09-729-835-80	Sequence 80, Appl
3	418	2.9	4019	US-09-738-973-425	Sequence 425, App
4	416.5	2.9	2665	US-09-864-761-34248	Sequence 34248, A
5	403	2.8	3256	US-09-919-172-98	Sequence 98, Appl
6	401.5	2.8	1367	US-09-801-368-108	Sequence 108, App
7	401	2.7	5179	US-09-815-242-12713	Sequence 12713, A
8	391	2.7	5179	US-09-922-217-1068	Sequence 1068, Ap
9	391	2.7	5179	US-09-833-263-1068	Sequence 1068, Ap
10	355.5	2.5	2478	US-09-815-242-5816	Sequence 5816, Ap
11	355.5	2.5	2478	US-09-815-242-12967	Sequence 12967, A
12	336	2.3	2843	US-08-681-219-32	Sequence 32, Appl
13	323	2.3	1596	US-09-902-432-4	Sequence 4, Appl
14	321.5	2.2	1400	US-09-764-176-7	Sequence 7, Appl
15	321	2.2	2063	US-09-735-367B-2	Sequence 2, Appl
16	320.5	2.2	1781	US-09-738-877-3	Sequence 3, Appl
17	318	2.2	2005	US-09-735-367B-3	Sequence 3, Appl
18	317	2.2	1363	US-10-124-557-52	Sequence 52, Appl
19	315	2.2	6281	US-09-815-242-12996	Sequence 12996, A

20	312.5	2.2	1140	US-10-124-557-104	Sequence 104, App
21	312.5	2.2	1404	US-10-124-557-2	Sequence 2, Appl
22	312.5	2.2	1404	US-10-124-557-62	Sequence 62, Appl
23	311.5	2.2	1049	US-10-124-557-58	Sequence 58, Appl
24	311.5	2.2	1313	US-10-124-557-142	Sequence 142, Appl
25	311.5	2.2	1354	US-10-124-557-48	Sequence 48, Appl
26	309.5	2.2	2368	US-09-815-242-5635	Sequence 5635, Ap
27	309.5	2.2	2368	US-09-815-242-12389	Sequence 12389, A
28	309	2.2	2441	US-10-109-886-8	Sequence 8, Appl
29	306	2.1	957	US-09-922-217-1065	Sequence 1065, Ap
30	306	2.1	957	US-09-833-263-1065	Sequence 1065, Ap
31	305	2.1	1320	US-10-124-557-46	Sequence 46, Appl
32	305	2.1	1320	US-10-124-557-60	Sequence 60, Appl
33	305	2.1	1361	US-10-124-557-40	Sequence 40, Appl
34	304	2.1	1314	US-10-124-557-50	Sequence 50, Appl
35	301	2.1	1038	US-10-124-557-74	Sequence 74, Appl
36	301	2.1	1270	US-10-124-557-44	Sequence 44, Appl
37	301	2.1	1311	US-10-124-557-42	Sequence 42, Appl
38	300	2.1	1022	US-10-124-557-84	Sequence 84, Appl
39	297.5	2.1	5795	US-09-815-242-12610	Sequence 12610, A
40	296	2.1	1537	US-09-801-368-104	Sequence 104, App
41	295	2.1	2442	US-10-109-886-10	Sequence 10, Appl
42	293.5	2.0	941	US-10-124-557-14	Sequence 14, Appl
43	289	2.0	1111	US-09-815-242-12955	Sequence 12955, A
44	288	2.0	1070	US-09-735-367B-6	Sequence 6, Appl
45	286	2.0	128	US-09-764-864-1571	Sequence 1571, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-816 Application US/09925297
Sequence 816, Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (178)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (269)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-297-816

Query Match	10.3%	Score 1474;	DB 10;	Length 328;
Best Local Similarity	94.8%	Pred. No. 9.1e-58;		
Matches 289;	Conservative 0;	Mismatches 14;	Indels 2;	Gaps 1;

Qy	406	NKSLAAANEELIESIRAKKGDIDINVP--EEFEKDKNLTENDSKAEKNREFEQSL	463
Db	12	NVLSLGGTNEELIESIRAKKGDIDINVPSPGEEFEKDKNLTENDSKAEKNREFEQSL	71
Qy	464	EKSDSDKTPDDDPQGSSEVGDFFKSEKSNELSSPACGAGAGSTYIIITRLNPNPSKUS	523
Db	72	EKSDSDKTPDDDPQGSSEVGDFFKSEKSNELSSPACGAGAGSTYIIITRLNPNPSKUS	131
Qy	524	QLKSOOYAAAHENANKLFKEGKEVLYVNSGGEISRLSTKKEVIMKGNINNYFKLGGEGKY	583
Db	132	QLKSOYAAAHENANKLFKEGKEVLYVNSGGEISRLSTKXXVVMKGXINNYXKLGGEGKY	191
Qy	584	RYVHNQYSTNSFALNKQHRDHDHKKRRHLAKFCULTPAGFEKNNGSHVSGKYLFTSRL	643
Db	192	RYVHNQYSTNSFALNKQHRDHDHKKRRHLAKFCULTPAGFEKNNGSHVSGKYLFTSRL	251
Qy	644	TITQLENNIPSSFLPNWASHRANWIKAVQMSKREFALALILECAVNPVYMLPIWRE	703
Db	252	TITQLENNIPSSFLPNWASHRANWIKAVQMSKXREFALALILECAVNPVYMLPIWRE	311
Qy	704	FLGHT 708	
Db	312	SLGHT 316	

RESULT 2
US-09-729-835-80
Case 9:09-cv-00080-MS Document 1-1 Filed 04/14/10 Page 1 of 1

```

? Patent No. US20010016647A1
?
? GENERAL INFORMATION:
?
? APPLICANT: Ruden et al.
? TITLE OF INVENTION: 29 Human Secreted Proteins
? FILE REFERENCE: P2015P1
? CURRENT APPLICATION NUMBER: US/09/729,835
? CURRENT FILING DATE: 2000-12-06
? PRIOR APPLICATION NUMBER: 09/257,179
? PRIOR FILING DATE: 1999-02-25
? PRIOR APPLICATION NUMBER: 60/056,270
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,271
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,247
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,073
? PRIOR FILING DATE: 1997-08-29
? NUMBER OF SEQ ID NOS: 128
? SOFTWARE: PatentIn Ver. 2.0
?
? SEQ ID NO 80
?
? LENGTH: 238
?
? TYPE: PRT
?
? ORGANISM: Homo sapiens
?
? US-09-729-835-80

```

Query Match	6.9%	Score 982.5	DB 10	Length 238
Best Local Similarity	78.8%	Pred. No. 1.5e-36		
Matches 189; Conservative	7	Mismatches 19	Indels 25	Gaps 3

Qy	2563	PLPLVASQRRKREER-----DSSSKSKKKKMIITTSKETK-----	DT	2601
Db	3	PTPR-----KTPYDEFSTYIGCDLCTMWYGECEGIIIEAKKMDYVTCNDCKRAQGSSE		58

Qy	2602	KTYCICKRPYDESKRYIGCDRCOMWGHGCVGLGSEALIEBYVCPQCSFEDAMTVLT		2661
Db	59	ELKICIRTPYDESQFIIGCDRCOMYHGKVCGLGSEALIEBYVCPQCSFEDAMTVLT		118

Qy	2662	PLTEKDEGLKRVLRSLQAHKMAAPLELPVDPNDADYGVIKPEMDLATMEERVORRY		2721
----	------	--	--	------

Db	119	PLTEKDYEGKLKRVLSQLDAKKMAMPELPEPDPADADYGVYKEPDLATMEERQVR	178
QY	2722	EKLTEFPADMTKLPDNCRRYNNPSPDSEPYQCAEYLESEFPQKLKGFKASRSHNNKLQSTAS	2783
Db	179	EKLTEFPADMTKLPDNCRRYNNPSPDSEPYQCAEYLESEFPQKLKGFKASRSHNNKLQSTAS	238

```

, RESULT 3
, US-09-738-973-425
, Sequence 425, Application US/09738973
, Patent No. US20020110563A1
, GENERAL INFORMATION:
, APPLICANT: Reed, Steven G.
, APPLICANT: Henderson, Robert A.
, APPLICANT: Lodes, Michael J.
, APPLICANT: Fling, Steven P.
, APPLICANT: Mohamath, Raodoh
, APPLICANT: Algate, Paul A.
, APPLICANT: Secrist, Heather
, APPLICANT: Indrias, Carol Yoseph
, APPLICANT: Benson, Darin R.
, APPLICANT: Elliot, Mark
, APPLICANT: Mannion, Jane
, APPLICANT: Kalos, Michael D.
, TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
, TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
, FILE REFERENCE: 210121.475C9
, CURRENT APPLICATION NUMBER: US/09/738,973
, CURRENT FILING DATE: 2000-12-14
, NUMBER OF SEQ ID NOS: 587
, SOFTWARE: FastSeq for Windows Version 3.0
, SEQ ID NO: 425
, LENGTH: 4019
, TYPE: prt
, ORGANISM: Homo sapiens
, US-09-738-973-425

```

Query Match	2.9%;	Score 418;	DB 10;	Length 4019;
Best Local Similarity	17.4%;	Pred.No. 1.2e-10;		
Matches 497;	Conservative 320;	Mismatches 995;	Indels 1018;	Gaps 103;

[illegible]

Qy	638	LPIRRELGHTRLRHMTSIREKEKVVKKKKOE- -ET- -----MOATWKTTFPVKH	751
Db	377	---WSEDLPTLVDESVS- -TESTEKKIKRYKRKKKLEETEPAYLOAFEGKDLIDYR	432
Qy	752	QVMQKGEEXRVYGVGWSMISKTHYRRVPLKPGTNNVYKKSLEGRKNNMDEMDS	811
Db	433	Q----- -SKTSDNLSDEGOADLYKTNN----- -TGFIDPSLD--P	465
Qy	812	KRKSRSPPKIKIETPDEKNEVGSAAKGAODNEMDISKITEKKDODVKELLDSDSKP	871
Db	466	LSSSSAPTK----- -SCTHGRADPLADISEVL- -NDDDLGIISDDLAKS	510
Qy	872	CKEEPM- -EVDDMKTESHVNCQESSQVDVYVNSEGFHLTYSKKTKKSKLDGLIERIK	930
Db	511	VHSDIDIPVDDPSSLPQPVNOS- ----- -PLSEQDGLI-	547
Qy	931	QFTLEEKQRIEIKLEBGJG- -IGRTSTNSKNLSPIYTAKKSCQSDSKROQSP-	987
Db	548	----- -SPELDKMYTDAILGKLYKIPELGKGDVELETAVLSPANTOPPLPQPPPTQ	601
Qy	988	----- -NANNDQEDLIGCQSQSDSVYLKMSDP----- -SHTTAKLYPKDVLDDVSTR	1034
Db	602	LPLIHNDQARSRLPMNGLLGSSPHLPHNSLPPGSLGTFMSLAQSYDPAR- -DKNSAF	659
Qy	1035	SEPFCKPQN----- -SIENDIEKVSDLASRQEPYKSKTKGDNFFIDDSKLASADDG	1088
Db	660	NPMASDNNNSMTSAPTVEGE- ----- -NDTMSNAQRSTLKWER----- -EALGEMATVA	707
Qy	1089	TJLICKN- -KKPLIOESTYIYSSSKSALHSVPKSTNDRAATPLSRAMPEGLGCD- -SE	1145
Db	708	PVLTTNINFPKLEEPDWTTRVKQ- -TAKIMRASSQERAPYQVQARJNRALRINKVM	766
Qy	1146	SNSTLENSSDTVSIOJDESE- ----- -EDMIVONSNEISISEQPTTRQO----- -DV	1187
Db	767	SNDMKRQOQDDSIDPSRIRLDESELFNDPLKQRSERH- -EGEMKFFRQOMQKSKQAKIEAT	825
Qy	1188	EVLPELCE----- ----- -LVSGSTGNCEDRLPVKCTEANGK- -KPSQO-----	1235
Db	826	QKLEQVKNEDQ00000QFGSOLHLVQSGDSPSSGISOPLTPQPGNGMNSPQSFKELF	885
Qy	1226	- -KKEEPRVKKCSQIKLMTT----- -DKNNKRESEKKGQRTSTQJNGK	1271
Db	886	TKQPPSTPTSTSDVYFKQOAPPPPPAPSRIPIDQSLSOAQTSOPSPQVFPSSSSMR	945
Qy	1272	DNKP- -KTYLK----- -GELKEISESRVYSGVPEPKVNNIKI	1307
Db	946	PPSPMDYAKMVGTPRRPPYGFHSFRNSNAAPENCOTPLSISYRPLQAN- ----- -ETTANRP	1001
Qy	1308	IPENDI- -KSLTYKESAIRPFGIDVIMEDFERNSEETKSHLLSSDAEG-----	1356
Db	1002	SPVBDLCSSSTNNDPYAKPPTPRPVYMDQFPKSLGSLRSPVSEBQAKGFIAGTSDH	1061
Qy	1357	----- -NRSLEFLPLSTKESDSTQTTTPSASCPE- ----- -NSYNQV-E	1393
Db	1062	FTKSPRADYFQORQIPDSYARPLR- -PAPLDSGPGFTPQPPPSQDPYGSYSQASR	1120
Qy	1394	DMETSEVKKYKTSPIITSEENLSNDFIDENGDLINKENVNQESKRRKTYITEVTMT	1453
Db	1121	RLSDYPERALRPLRPLDNSH- -MOSNDYSDQ- -PLTHPRVYN----- -ESFAHP	1167
Qy	1454	STVATESKTYIKVEGDKQTVASST- ----- -ENCAKSTVTTTTVYTKLSTPSTGGSVUI	1508
Db	1168	SRASQCGTISRPTSDQPYQSPGTPRPVVDYSQSGSGRARSNTDPIYSQPEGTPTTVD	1227
Qy	1509	SVKQSKTIVTTVTYDLSLTGTLVTSMATVSKESYTRKXVYLKFSRPKTRSTALPS	1566
Db	1228	PYSQOPPTPRPSQTQDLEFVPR- ----- -VTNORHSDPYVNHPRGTPRGISYVYSQPPATPRPR	1283
Qy	1569	YRKIVTSTKKSIFVLBNDDKLKLARKGIREVRYN- -YNNKPLALDIMPVSPSPRTGI	1626
Db	1284	ISEGFTSSMTRPVLMNOD- ----- -PFLQAAQNNGPAL- -PGGLVP- -----	1333
Qy	1627	TWRRLQTVSLAGVSLMRLMLASLRWMDMAKVPGGGSTRTETSETEITTEIIKR	1666

[illegible]


```

Db 2241 -ETQOLLQMOQKYLEQIGAHRSKRLSAKOR-----T 2273
QY 2451 AKRKRRESEVQOKRSKONATKLSALLFKHKOLRAELIKKALLDKDLOJEVOELKRD 2510
Db 2274 AKRAGRFEPEDEAQLKHTVEQOSWQKOLEQIRK-----QCKEHAELIEDY 2320
QY 2511 KIKKEDLMQLOATAVAAD-CPVPTPVLPAPRAPPPPPPGV-----QHTGLLST 2562
Db 2321 RIKQOQ---OCAMAPPTMMSVQOPPLIPGATPPTMSQCTFFPMVPOQLQHQOHTVVISG 2377
QY 2563 PTLPV 2567
Db 2378 HTSPV 2382

RESULT 4
US-09-864-761-34248
Sequence 34248, Application US/09864761
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-x-1
CURRENT APPLICATION NUMBER: US/09/864, 761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180, 312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207, 456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632, 366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236, 359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234, 687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608, 408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774, 203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34248
LENGTH: 2665
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL034555.2

```

```

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
OTHER INFORMATION: EST_HUMAN HIT: A0117052.1, EVALU0 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P08640, EVALU0 3.00e-10
US-09-864-761-34248

Query Match 2.9% Score 416.5; DA 10; Length 2665;
Best Local Similarity 18.5% Pred. No. 8.9e-11;
Matches 544; Conservative 422; Mismatches 1145; Indels 825; Gaps 132;

QY 7 EEEEDGAETQDESEDE-----EDEMEDDDSDYPEEMEDDDDAASYCTESSFRSHS 59
Db 170 KNEKTKERTFDEPERERERRLIRKEVEKDKTDQ----- 205
QY 60 TYSTGRRKRPVHRPRSPLE-----EKDIPLEPKS-----SEDLMPVNEHI 104
Db 206 -----KRGKGVHSPSSQSESTDQENEREQSP-EKPRSCNKLRSRKADEKGIATKRLRE 257
QY 105 MNVAIYEVLRNFGVLRSLSPRPFEDFCALVQEOCTTMAENHVVLLKAVLEEDTSNT 164
Db 258 MCPVVLTRVKEKEGKVIDHTPV-----EKIKALNDIYKSSALDQKLOYSP 305
QY 165 TFGPADLK--DSVNSTLYFLDM--TWPEVLRVYCESD--KEYHNVLRYQEAEDYRGY 218
Db 306 EPAKSDLSKLESVRYMKVPKGLSSHYEV--EKEGRKAKRHILKPEQPAD--GVS 357
QY 219 ENKIKVLOFLVDQPLTNTIARELMSE--GVLYQYDHCYVCKLGDLLCCETCSAYHL 275
Db 358 AVDLKLEARKRRFADSNLAKQKQEVKKSSPEMD-ARVLSKKQPDVSSREVIILRG 416
QY 276 EGVKPPV--EEVPEDEMOCEVCYAHKVPGVTDQVAETQKKRPYIRHEPIGYDSSRRKYWF 333
Db 417 EAERKVPYKRIILKRESKIKLIDLNTYASPKDC-QELAS-----ISVSGSGSPSSD 466
QY 334 LNRRL--IIEEDTENENEKKIWIYSTKVOIAELIDCLDKYWEAECLKILEEN--REET 388
Db 467 LQARLELAGSEVQENQVQSKRPISPKQLQK-----QVLDQGGPEREDV 512
QY 389 HR-HMDITEDLTAKAGSNKSPFLAANELLSEIRAKG-DIDMNVSS--PEETEDKNFE 443
Db 513 RKNVCSLRDETPEKSGQEKSHSVNTEKI-----GIDIDTQSYRKQMEQSRKQ 564
QY 444 TENDSKDAE-----KNREFEEDQSL-----EKSDDKTPDDDEQCK-SEVGD 486
Db 565 MEMEIKASEKFGSPKDVDEYERKSLVHEVGKRPQVDTDSPSKKRMADHYDFDICTR 624
QY 487 -KSEKNGELSE-----SPGAGKA-----SGSTRITRLRNP--DSKLSQKSOQ 529
Db 625 ERNVRSSRQISEDSERTGSPSVRHGSHFHEDEDPISRLSVKSGPKYDEKVLPSNIT 684
QY 530 YAAAHAENKLFKEGKEVL-----VNSGELISRLST-----K 562
Db 685 VRESLEKFNPDYSSRQOMADMAKIKLVLNSEDENRMDQSKODAGRDVSPFNSIIR 744
QY 563 KEVIMKGNINNY-----FLGQEGRYVYHNGYSTNSFLNKH-----QHRD 605
Db 745 RDSLKRKRVRLDEPGEVPSDSEDEG-----HKSHSPRASLSYSSRLSLRLRREKLAE 800
QY 606 HDKR-----RHIAHKEFCL--TPAGEFKWNSGVSGVLTISTRLTITQLENNIPSS 655
Db 801 RDERLSSLSERNKFYSFALDKTTP-----DTKALLERAKSSSSREN--WS 846
QY 656 FLHPNMASHRAKAWIKAVQMSKPRERFALALIECAVKKPVVMPIYREFLGHTRLHRMYS 715
Db 847 FL--DWDSRFANFRN-----NKDKE-----KVSADAPRPI--PSW--YMKKKKIR 884

```


GENERAL INFORMATION:
APPLICANT: Paris, Mary
APPLICANT: Turner, Christopher M.
TITLE OF INVENTION: PROSTATE CANCER MARKERS
FILE REFERENCE: PA-0036 US
CURRENT APPLICATION NUMBER: US/09/919,112
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/222,469
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PERL Program
SEQ ID NO 98
LENGTH: 3256
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

Query Match 2.88; Score 403; DB 10; Length 3256;
Best Local Similarity 18.5%; Pred. No. 4.3e-10;
Matches 523; Conservative 371; Mismatches 1011; Indels 916; Gaps 130;

QY 339 IIEEDTENENKRIWYSTKVQALIDCLDK---DYWEALCKILEEMREETHIRMDI 394
DB 432 VLPETEIHNEPFLTLMTLOVERKIOKDSLSKPEKLGTTAGOMSGILGLSS-----VDI 486
QY 395 TE--DITNKARG-----SNKSFLAANELE-----SIRAKGDIIDN-----VKSPEE 436
DB 487 NNGSGINESGIPLKRIRRVSGFGLHRLPELDENLPPNTPLKRGAEAPTRKRSKSLVMHTPPV 546
QY 437 TEKDKNEFENSKAENKRE---EFEDQSL---EKSDDKTP---DDDEQCKSVGDFKS 488
DB 547 IKTIIKEPOPSPGKQESSEIHEVKAOSLVISPPADSPRTPAVASDORRSCCTAASS 606
QY 489 EKSNGELSESPGAGKAGSTRITRLNPNDS-----KLSQLKSOQVAAAHBANKLFK 542
DB 607 SKSOFEVK-----RGERVATCLOKRVYSISQOHDLQMCISKRSGASEAN-----654
QY 543 EGKEVLLVNSOGELSRSTKR---EVIMKG---NINNYFKLGQCKRYV---YHNOYSTN- 593
DB 655 ---LIVAKSVADYKILGAKOTQTKVIKHGPORSMNKRORRPAATPKRPGVEVHSGFSTGH 710
QY 594 ---SFLANKHGHEDDKRR---HLAKHFCULTPAGEFK-----WMSVYSGSKYLT 637
DB 711 ANSPCTIITIGRAHTEKHAVPAPRYVNLNFI SNOKMDKEDLSGIAEMFKTPVQEPOLT 770
QY 638 ISTERLITTOLENNIPSSF-----LHPNMAASHRAMIKAVOMCSK--PREFALALAI 687
DB 771 -STCHIAISNSENILGQFOGDSGEPLPTSESFGNVEFFSAONAKOPSD-----822
QY 688 LECAYKPVVMLPIREFLGH-----RLHRTMSIRE-----719
DB 823 -KCSASP---PLRQOCIRENGNNAKTBRNTYKMTSLTKISDTEPESKTVSVNRSGR 877
QY 720 --EKKVKK--KEKKOEEETM-----QOATWVRYTFPVKHQVKKOGEEERYVGY 766
DB 878 STEFNRICKLPVESSEENIEIVICILKRGOKATLLOO-----RRGEMKEI---925
QY 767 GGSWMSIKTHYRVPKLPGNTNVYRKSLSGTKNNMDENDESK-----RKCS--RSP 819
DB 926 -----ERPFET---YKENIELKEN--DEKKAMKRSRTWQKCAPMSDL 964
QY 820 KRIKTEPSE--KDEYKG-----SDAAKGADONEMDISKITEKKDOVVKELDSDSKP 871
DB 965 TDLKSLPTELKDTARQONLLOTODHAKAPKSEKGIYK-----P 1006
QY 872 CKE-EPMEVDMDKTESHNCQESSQVAVNVSE-----GPHLRYS-----911
DB 1007 CQSLQP-----EPINTPTHTKQOLKASISGVYKEILLAVGKTRTSGETHTHREPAGD 1062
QY 912 -----YKKKTK-----SSKLDGLLEBRRIKOPTLEEKORLEKIKLGGIKIGIKTSTNSK 961

DB 1063 KSLRPFKESPKQIILDPAAAVTGM--KKMRPTEKEBAOSLEDL---AGFEKELFQTPGSPSE 1117
QY 962 NISESEVITIKAEKGOSDSMRQDSNANNDQPE--DLIOGCSQSDSVLRMSDPHHTNK 1020
DB 1118 SMDEKT---TKIACKSPPESDVPTSTKOWPKSLRKAADVEEELAKRLTPSAGRAM 1174
QY 1021 LVPKDRVLDDVSI-----RSPETCKCPKONSIEH-----DIE 1051
DB 1175 LTFKPAGGBEKDIKAMGTPVQKLDLAGLTPSKRROLQPRKKAQALBDLAFKELFQTP 1234
QY 1052 EKVSDDLASRGQ-----EPYKSKTKGNDFTIDSKLASADDTITL--CKNK 1095
DB 1235 GHTEELVAAAGTKTKIPCDSPQSDPVDTPSTKOR-----PKRSIRKADVECELLACRNL 1288
QY 1096 KPLIOESDITVSSKSAIHSSVPKSTNDR-----APFLSRAMPEGKLGDSSENSYL 1150
DB 1289 MP-----SAGKAMHTPKPSVGEERDIIIEVGTIVQKL-----DLT 1323
QY 1151 ENSDITVSIODSSSEDMIVQNSNESISEQFRPREQDVEVLEPLKELVSGESTG--NCEDR 1209
DB 1324 ENITGSKRRRQPTKEEAQALBDLTFKELFQTPGHTEEA-----VAAGKTKMKPCESS 1376
QY 1210 LPVKGTEANGKPKSQCKLEERPYNKCSQOIKLAKNTTKKNENRSEKKGGRTSTFOIN 1269
DB 1377 PPEADPTPTSTRQPKPTPEKRDVQKELSAK-----KLTQTSGETHTTDKVP 1424
QY 1270 GKNNKPRILYKGCLEKIEISRVSGNV--EPKVNINKIIPENDISLTVKESAIRFTN 1328
DB 1425 GGECKSINARETAKOKLDPAAVYTSKRRPKTK--EKQAPLEDLAGW--KELFQTP---1477
QY 1329 GDVIMEDFENRNSSETKSHLLSSDAEGNYRDSLETLPTSKESDSQTOTTPSASCEENS 1388
DB 1478 ---VCTDKPTTHEKTKIACRSQPD-----PVDTPSSKQSKR 1513
QY 1389 VNOVEDMEITSEYKKTYS-----PTSEESNLNDFTIDENGLPK--NEWV 1436
DB 1514 SLRWVDSEEEFALRKRTPKAGRAMHTPKPAVGEKIYA--FM---GTPVQKLDLEML 1568
QY 1437 NGEKRTVITTEVTTMTSTVA-----TESKTVIKVKGDKQTVVSTENCANS 1484
DB 1569 TGSRRRLQPTKEKAQALBDLAGFEKELFQTRGHTEESMT---NDKTAACKSSQOLDKN 1625
QY 1485 TVTTTTTVTKLSTPSTGSGVDIISVKEQSKYVTTTVDSLTGTLVTSMTVKEYS 1544
DB 1626 PASSKRLKTSIG--KGVGVEELIAGKLOTQSGETHTEPTGOKSKRAKMESPKOT 1683
QY 1545 TROKVKLMKTSRPKKTSGTA-----LPSYRK--FVTKSTKSIFFVLPNDD 1588
DB 1684 LDSAASLTGSKROLRTKKGSEVPEDLAGFELFQTPSHTKESMTNKTTKVGYRASQPD 1743
QY 1589 L-----KTLARKGCIREVYPYNNAKPALDITWYPSPPRFGITWRYRLQTVKS 1637
DB 1744 LVDTPTSSKQPKRSLKADTEE-----EFLARQKQTPSAG---KAMHTPKP 1787
QY 1638 LAGVSLMR--LIMASLWMDMAAVPPGGSTRTETSETETITTEILKRRDVPYGR--1694
DB 1788 AVGEKDNITFLGPPVKKLDPGNLP--GSNRRLOTKREKAQALBEIT---GREL 1838
QY 1695 -----FEYCIRKIIC--PIGVP--ETPKETPPQKGLSS-----ALRKP 1733
DB 1839 FQTECTDNPTTDEKTKKILCKSPQSDPADTPNTKQPRKSLKADVEEFLAFKRLTP 1898
QY 1734 -----ETPKQTPGVIIETWVAEELELMEITRAF---AERVE-----KE 1768
DB 1899 SAGKAMHTPKAA-----VGEER-----DINTFVGTPVKKLDLGLNLPKSKRRPQTPKE 1946
QY 1769 KAAQAVEQAKRRLDOQPTVATATSTPSSTSTISIPACKVWVAPISGAVTTGKAVLT 1828
DB 1947 KAKALEIDLAGEKELFQTPGHTSEESMTDKTIEVSCSKSPQDPVPTP-----TSSKRLK 2000
QY 1829 TKVSPATVTPQOKKNFHQFPATWVWQOQNSGVQVQOKVL--GIISSSTGTSQGTFTS 1886

2001 ISLGR-----VGKKEEVLPGKLTQTSKTTQT--- 2028
QY 1887 FOPRATVTRPNTSGSGT-----TSNSQVI-----TGQIRPGMTVTRP-----L 1929
Db 2029 -----HRETADGGSIAFKESAKOMLDPANYGGMEMW-----RTKEKAQSL 2073
QY 1930 QOSTLGAIRPVPVQAPAPQOVMTQIIRGQVSTAVSAPNTVSTPGOKSLTSATSTS 1989
Db 2074 EDLAGFELFQTPDHTEESTDDKTKI-----ACKSPPESSMDPTST- 2117
QY 1990 NIOSSASOPRRQOGQVKT--MAOLTOITQHGNGGLTVVIOGOGQITGOLQILP--- 2044
Db 2118 -----RRRPKTPPLGKRDIIVEELSLALKQLTQ-----TTHTDKVPGE 2153
QY 2045 -OGVTVLPFGQOIM-----QAAMPNGTVQ-----RFLV--TPLATVATTAS 2083
Db 2154 DGINFRETAKOKLDPASVYSGSKRQPRTPKCKAKOPLDLAGLKFOTPICTDPTTH 2213
QY 2084 TTTTVVSTTA-----AGTGE--QROSKLS--POMOVHODKTLPPASQSSVGPANAPOTA- 2134
Db 2214 ERTTKIACRSPQDPDVGTPTEIFEPQSKRSLRKADVEEESIALRKRPVSGKAMDTPKAG 2273
QY 2135 -----QPSARPOQOTQPSAPQPEVQOTQEVQOTVYSS-----HVP 2171
Db 2274 GBEKDKAMGTBPVOKLPLGNLPGSKRWPQTPKE-----KAQLEDLAGFELFQTP 2326
QY 2172 SEAPRTHAASSKPOVAASQPOSNVQSGSPVRSQSPTRIRPS----- 2215
Db 2327 GMDKPT--TDEKTTKIACKS--PQ-----PDVDTPTASTKQKRRKRLRKADVEEFLALR 2377
QY 2216 --TPS-----QLSPGQOSQOVOTTSQPIPIQPIHTSL-QIP--SOGQPOGSP----- 2256
Db 2378 KRTPSAGKAMDTPKPAVSEKNNINTEFETPVOKLIDLGNLPGSKRQRPQRPKEKALEDL 2437
QY 2257 -----QVQSSQOTLSSQOTLQNVSSPSRQOIQIQBPQVIVP-----QLQOQVQY 2305
Db 2438 VGEKELFQTPGHTHEESMTDKITEVSCSKPQPSFETSSSKRKLKPLPVKMDKEEPLA 2497
QY 2306 LSGQI--QSOVNAQIOAQSGVPOQIKL--QLPIQIOSSAVQTHQIONVTVQAASVOEQ 2361
Db 2498 VSKLTPTSETTQTHTEPFGDSKSIKAFKESRQOILDPAA-----SVTGSRRQLRTR 2549
QY 2362 LORVQOLRQOQOKKQOQIETKREHTLQASNOSEI-----IO 2398
Db 2550 KEKARALEDLVDFKELFSAPGHTHEESMTIDKNTKIPCKSPPELTPTATSTKRCPKTRLR 2609
QY 2399 KOVVMHNAVIEHLKOKKGMT-----PAERENQMIYCNQVMKTIILDKIDKEEQAK 2452
Db 2610 KEVKEELSAV--ERTJOTSGQSTHTHKEPASGDEGIVL-----KORAK 2651
QY 2453 KKRREESVEQKRSKONA-----TKLSALLFKHKKEQLRA----- 2485
Db 2652 KKPNPVEEPPRRRRPAPREKAKOPLDLAGFTLSTSGHTQESLTAAGATKIPCESPPL 2711
QY 2486 ELIKKRALLDKLOLEVOBELKRDLIKKEKDLMLQAO----- 2523
Db 2712 EYVDVTAASKRHLRTFYOK-----VOVKEEPSAVKFTQTSGETTDADKEPAGEDGIAL 2766
QY 2524 ---ATAVAPCPVPVPLVLAAPAPRPPSP-----PPRGVQHTGLSTPIPLVASQKR 2572
Db 2767 KKSAAQTPAPASVYSGRRRRPAPRESAOAIEDLAGFKDPAAGHTHEESMTDOKTKIPCK 2826
QY 2573 KKEEEKSSSKSKK-----KMISTSKETKRPYKLYCICKTPYDES 2614
Db 2827 SSPELEDFTSSKRRRRTRAQKVEVKEELLAVGKLTQTSGETTHND-----KEFVGE 2879
QY 2615 K 2615
Db 2880 K 2880

RESULT 6
US-09-801-368-108

Sequence 108, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 108
LENGTH: 1367
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Best local Similarity 2.8%; Score 401.5; DB 10; Length 1367;
Matches 289; Conservative 168; Mismatches 564; Indels 339; Gaps 56;
QY 1135 DFEQ---KIGCDE-----SNSTLENSDPTVSIOSSSEDMIVQNSIESIQFRT 1182
Db 191 DFGFVWINIDCNNCGTKSTSTTSSTSESTTSSTSESS-----TTSSTSEST 243
QY 1183 REDDVEYLELKEKELVSGESTGNC--EDRLPVKGTENGKRPQOKKLEERPYNKSCDOI 1240
Db 244 TSSSTSESSSTSTTAPATPTTCTCKEKPPTPTTSCYKPKPP-----PHHDTPTCT 296
QY 1241 KLNRTDKKNENRESKKGORTFOINGKDNKPKYILGECIKETESRVSQVNEPK 1300
Db 297 KKKTTTSKTQTKTTTPVPSSSTTB-----SSAPVPPSSSTTESSAPVSSSTTB-- 350
QY 1301 VNNINKIIPENDIKSLVKSARPTPLNGDVIMEDFERNSSSTKSHLSSDAEGNYRD 1360
Db 351 ---SSAPVP--TPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY 1361 SLETLPTSTKSDS---TQTTTPASCPESNSVQVEDMELETVEKKVYSSPTTSE---- 1413
Db 395 VPPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPV--TSSTTESSAPVTSSTTES 453
QY 1414 -----EESN--LSNDFIDENGLPINKNENVNGESRKRYITEVTMTSTP--VA 1457
Db 454 SSAPVPPTSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVP 513
QY 1458 TESKTIYKEKQKQTVVSTENCASVTYTTTQTLSTPSTGSGVDIISVKEOSKTV 1517
Db 514 TPSSSTTESSAPAPPTSSSTTESSAPVTSSTTESSAPVPTSS--STTESSSTP 568
QY 1518 VTTTVVDSLTGTLVTSMTVSK--YSTRDKVYKLMKSRPKTKTSAGTALBPYKRFYK 1575
Db 569 VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS 628
QY 1576 STKKSIFVLVNDLKLKARKGIREVYFMYNAKPALDIMPYSPPTFGITRYRLQVY 1635
Db 629 STTESSA-----PVTPSSS-----TTE 647
QY 1636 KSLAGVSLMLRLMLASLRLMDDMAKAVPPGGSGSTRTETSETEITP-----TEIKR 1685


```
Db 648 SSSAPVPTP-----SSSTTSSSAPVPTPSSST-TESSSAPVTSSTESSAPVTSSTTE 701
Qy 1686 RDGPFYIREYELCKRTKICIVP-----ETPKETPTPOKGLRSSAL----- 1728
Db 702 SSSAPVPTPSSSTESSSAPVPTPSSSTESSSAPVPTPSSSTESSSAPVTSSTESS 761
Qy 1729 -----RKRPEPQKOTGVIIETWAELELMEIRAFAEKREKKAQVDAQKRLLEQ 1784
Db 762 APVPTPSSSTESSSAPVPTPSSSTESSSAPVPTPSSSTESSSAPVPTPSSSNTSS 821
Qy 1785 KP--TVIATSTTS-----PTSSTSTISPAOKVMVAPISGVYTGKMLTKRVS-- 1833
Db 822 APSSTPSSSTESSSVVPTPSSSTESS-----APVSSSTESSVAPVPTPSSSSN 874
Qy 1834 -----PATVTFQOKNKHOFATWVKOGQNSGVVQYQKVLG-----IIPS----- 1875
Db 875 ITSSAPSSIPSSSTESSFST-GITVTPSSSKYPGSGTETSSTETTTIIVPTKTTTSVT 933
Qy 1876 -----STGTSQ--OTFTSFOPRTATVTIRPNTSGSGGTTNSOYITGPOIRPM 1922
Db 934 PSTTTTTCSTGTSNAGETTSKPKYVTTV-PTTTTTSVTSSSTTTT-----T 985
Qy 1923 TVIRTPLO--QSTLG--KAIIRT-PVMVQPGAPQOQVMTQIIRGQVSTAVS-----AP 1970
Db 986 TVCSTGNSAGETTSKPKITTTVPCSTSPSETASESTTSPTPTVTVSTTVTTE 1045
Qy 1971 NTVSTGQOKLTSATSTSNQSS--ASQPPRPOGOV-KITMAQLQLOUGHGNOGLT 2027
Db 1046 YSTSTPKPGE-ITTTFTYTKNIPTYLTIAPTPSTVTVTNPTTTTTCSTGTSNA-- 1102
Qy 2028 VVIGOGGOTTGLOLIPQGVVL-----PGPGQLOMQAMPGVQREFLTPLATATAS 2083
Db 1103 -----GETISGCS--KRYVTTTPCSTGTCGEYTTFA-----ITLVITTA----- 1138
Qy 2084 TTTTTVSTTAAGTEGROSKLSIPQMYHQDKTLPPAQSSSVPAKAPQTAQPSARPOQ 2143
Db 1139 -VTTVTVTTESTGTSNAGKTTTG--YTKSVPTTYVTTLAPS----- 1178
Qy 2144 TOPSPAPQEVOTQEVOTQTVTSNHPSEKQPHAOSSKRVQVAAQSPQSNQOGQSVR 2203
Db 1179 -----APVTPATNAVPTTTTECSAATINAGETTSVCSAKTIVSASGENTAPSAT-- 1231
Qy 2204 VQSPQTRIRPST--PSQLSPQOQSOVQT--TSQPIPIQPHTSLOIPSGQPOSOPOV 2258
Db 1232 --FTVTTAI-PTVYITTESSGTSNAGETTTGYTTKSIPTYITTL-IPGNGAKNNEYV 1287
Qy 2259 QSSQTQLSSGQTLNOVSVSPSPQLOIQOPQPOVIAVQLOQOVYVLSQIOSOVAAQIQ 2318
Db 1288 ATAT-----NPISIKTTS-----QIATTAASASSVAVV 1315
Qy 2319 AQQSGVPOQIKLOLPQIQOSSAVOHOIONV-VTVQAS 2357
Db 1316 TSPS-----LTGHLQASGSAVAITTSVPSSISTYQGA 1348

RESULT 7
US-09-815-242-12713
: Sequence 12713, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: EILTRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
```

```
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12713
: LENGTH: 2344
: ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 2.8%; Score 401; DB 10; Length 2344;
Best Local Similarity 18.0%; Pred. No. 3,7e-10;
Matches 336; Conservative 271; Mismatches 836; Indels 424; Gaps 57;

Qy 797 EGRKNNDENMDESDKRCRSRPRKTIKEPDSEKDEYKSGDAKADQNDMDISKITEKK 856
Db 747 EVIRKNSDVSSTGTSQOSVSTSKADSOASTSTSGS----- 786
Qy 857 DQVKEHLSDSDPKCKEPEMEYDDDKTESHYNCOSSQV-----DVNVSEGFHLRT 910
Db 787 -----IMTSTASSTKSTSVLSLDSVASKSLSTSESNVSSSTSTSLVN----- 831
Qy 911 SYKKTKTSKLDILER--RIKQFTLEEKQRLKIKIEGKIKIGTSTSSKNLSESPV 968
Db 832 ---SQSVSSMSGSVSSTSLSDF--ISNSSSTESK-----SVSTSTSDLRSTIS 878
Qy 969 ITRAKKGCQSDSMRQOPQPNANNQDPEDLIQCSQSDSVYLKMSDPSHTTNKLYPPDRVL 1028
Db 879 LSDSVMSSTGSLSKSGLSTSTSDASSTQSVSDSTSNBI-----STSESLSESGTS 932
Qy 1029 DVYSIRSEPTKCPKQNSIENDIEKVDLASRQGEPTKTKGNDFFIDSKLAS----- 1083
Db 933 ESISI-----SNSISNSVASTSKLESQSTISLSTIS-----DSKMSSTESL 975
Qy 1084 ADDIGTILCKNKKPLIGESDTIVSSSKKALHSVPKSTINDRQATP--LSRAMDEGKLG 1141
Db 976 SDSTST-----SDSVSGSLSVAGOSQVSTSDSMSTSEMIDSMSTSGSLA 1022
Qy 1142 CD-----SESNSTLENSDPTVSIQDSSEEDMTIVONSNEISBQFRTREQDVE 1188
Db 1023 ASDSKMSVSSSMSTSGSGSTSESLSDSISSTSDSKSLSLSTSQSGSTSTSTSSSVR 1082
Qy 1189 VLEPLKCELVGSESTGNCEDRLPYKGTENGKPKFSQOKLLEPRVNCQSPQIKLNTDK 1248
Db 1083 MSE---SQSTSGSMSTSQSDSTIS-----TSFSD-----STSDS 1114
Qy 1249 KNNENRESEKKGQRTSFQINGKDKPKIYLKGECLKEISRRVSGVGNVPPKVINIKII 1308
Db 1115 KSASTASSESTISQSVST-----STSGSVSTSLSTLSNSERTS 1152
Qy 1309 PE-NDIKSLTYKES-----AIRPFLNG-----DVIMEDFENRNSSETK-----S 1346
Db 1153 TSMDSSTSLSTESDSTSDSTSDSISEALISGSESTISLSESNSTSDSESKASAFLS 1212
Qy 1347 HLSSDAE-----GNVRSLETLDPSTKESDSTQTTTTSACSPESNSVQVEDMIEY 1399
Db 1213 ESLSESTSESTSESLSTSDSTSLSDSNSSESGSTSL--SNSYSGASISTSTSGSAST 1271
Qy 1400 SEVAKVTSSPTTSEENLSMDFFIDENGLPINKENNGE-----SKRKT 1444
```



```
Db 1272 STYK---SESVSTSLSTSTSLSTSLSTSLSTSLSTSGSKSNLSASMSSTSDSISTRKS 1328
Qy 1445 VITEVTTMTSTVATESK---TVIKVEKGDK-----QVVSSTENCAKSTVT- 1487
Db 1329 ESLASSTSLSGSTSESESGSTSSSEKSDSTSMLSMSOSTSGSTSVSTSESLSDSTSR 1388
Qy 1488 -----TTTTVTKLSTPSTGSGVDIIIVKEQSKTYVTTVTDS 1525
Db 1389 LSLASMSNOGVDNSASASASTSTSTSESDSTSSYSTSQSTSQSESTSTSLSDS 1448
Qy 1526 L-----TTTGGTLVMTVSKREYSTRDKYKLMKFSRPKTRSGTALPSYRKPFYKSTKKS 1580
Db 1449 TSLSTSGSGSTSTASISGSESESDSOSTSTSEKSESTSTSLSTSTSGNSGAS 1508
Qy 1581 IFVLPPNDLKLARKGIREVYFVYNAKPAIDIMYPSPRPFTGITWRYRLQTVKSLAG 1640
Db 1509 TSTL-----LSNSASASESDSSSTSLSDSTASMSQSESDS-----QSTSTSLN 1553
Qy 1641 VSLMLRLMASLRMDMAKVPFGGSGSTFTETSEITTEIIRKRDVGPYIGREYCIJR 1700
Db 1554 SOSTSTSMSTIASSEVSESTSESTSESTSESTSTSL----- 1595
Qy 1701 KIIICPIGVETPEKPTPTQKRLRSALRKRPERPKQGPVIEIETWVAEELMEIRA 1760
Db 1596 -----SDSOSTSRSTASGASASTSTSDSRSTASSTSMKRTSLD----- 1637
Qy 1761 FAERYEKEKAQAVEQAKRLEQOKPTVIATSTTSPSTSTISPAQKVMVAPISGYT 1820
Db 1638 -----SQMSLSTSTSTSVSDSTSLSDVSDSTSDSTSTSGMSASISLSDST 1688
Qy 1821 TGTKM--VITTVGSPAITYPOONKNEHOTFAFWAKOGOSNGVVOYQKVLGIPSSFG 1878
Db 1689 TSTASAEVNASISDSQMSSEVNSSEVSESNESDSKMSSTSVSS--GSLSVST 1746
Qy 1879 TSOQFTSTOPTATYTIPTNSGSGTTSNSOVIINGPIRPGMYIRPL--OQSTIG 1935
Db 1747 LRKSESYS--ESTISGSGMSDSVSTSDSSLSYSTQKRSSEVSESDSLSDSKSTSG 1803
Qy 1936 KAIIRTPVWVGAPQOVWQIIRQOPVSTAVSAPNTVS-----STPGKSLTSATSTEN 1990
Db 1804 STSTSTSGSLSTSLSGSESVSESSLSLSDSTSDSDSLSGSLSGSLST 1863
Qy 1991 IOSSASQPPRPOGOVKLTMAOLITOLTOGHGNGCL--TVYIOGOGTTOLOLIPGCVTV 2049
Db 1864 SDS-----LSDSKSLSSQSGSESTSTSVSDSOSTSGSTNSGPFDSMSISA 1909
Qy 2050 LRGPCQQLQAMAMPNGTVORFLFTPLATATTAATTTTAVSTTAAGTGROKSLSPQO 2109
Db 1910 SESDSKSTSDSSISGS-----NSTSTSLSTSDMSGSVSTSTSLSDSISGIST 1960
Qy 2110 VHODKTLPPAOGSSVGPAPAKAQPOTAQPSARPQOTOPQSPAPQEPVQTOPEVQOTVSSH 2169
Db 1961 VSDSS--TSTSESLSDMSAQOSTSTASGSLST-----SISLSMASAGILITSQST 2012
Qy 2170 VSEKAPHTAQSCKPOVAQOSPOSVQOGSPYRVOSPQOTRIRPSTPQSLSPGQOSVO 2229
Db 2013 VSTSLSTSDSIDSTISIT--SGSQSAVESES-----TSDSTISIDSESLSTSGSTSSST 2066
Qy 2230 TTTTSPPIPIPTSLQIPQOGPOQOPQVQSSQTLSSQTLINQVSVSPSRPOLQIQP 2289
Db 2067 TSTSE-----SLSTNSG-----STVSESLSTSGSGST-----SVDSSTSSLSST 2110
Qy 2290 QPOVIAPVPOLOQOVVLSQIOQOVAOIOAQOSGVPQOIKLOLPQIOQSSAVQTHQION 2349
Db 2111 GSTVSDSTSMESMSASISMSGOSTSGSTSTSTSESLMSGSTHSTSVS--DSOS 2168
Qy 2350 VYTVAAASVQEOLOK-----VQOLRDQOKKKQOOQIEIKREHTLOASNOSEIIOKOV 2401
Db 2169 ISTNSGMSNSIRHFTSLSTSLGMLSLSDSNMSDSDSVISIASSEMSMSMSDSVMSDS 2228
Qy 2402 VKK--HNAVIEHLKQKSKMTPERENQMIYCNQVMKILDKIDKEEQAOKKRKREES 2459
Db 2229 TSSSMNSMSMSTSESNLHPSNMSSTSO-----THFTSTSTSEISA-----PN 2274
```

```
Qy 2460 VEQKRSKONATKLALLFKH-----KEQL--RAELIKRALLDKLOIEVOELKRD 2510
Db 2275 TNSQSTLSATSVSS--KHDAEPAQSEERLPDTGDSINQNGLLGVMILLVGLG---M 2328
Qy 2511 KIKKEND 2517
Db 2329 KRRKKMD 2335

RESULT 8
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.47JC13
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match 2.7%; Score 391; DB 10; Length 5179;
Best Local Similarity 22.5%; Pred. No. 2.4e-09;
Matches 221; Conservative 91; Mismatches 418; Indels 254; Gaps 33;

Qy 1366 PSTKESDSTQTTTPASCEBSNVQVEDMEIETSEVKKYTSSPITSSEESNLNDPIDE 1425
Db 3460 PTTTPTTTTTPVPT--PTPTGOTPTTPTTPTTPTTPTTPTPTPTGOTPTT----- 3508
Qy 1426 NGPLINKENV-----NGESKRKTIVITEVTMTST--VATESKTVIAKVEGDKQTVV 1475
Db 3509 --PTPTTTTTPPTPTPTGOTPTTPTTPTTPTTPTTPTPTPTGOTPTTPTTPTTPTPT 3566
Qy 1476 SSENKAKSTVTTTTVTKLSPTSGSVSDIISVKEQSKTYVTTVTSLTGGT--L 1533
Db 3567 PPTPTGOTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTPT 3626
Qy 1534 VTSMTVSKREYTRDKYKLMKFSRPKTRSGTALPSYRKPFYKSTKKSIFVLPPNDLKLIA 1593
Db 3627 TTTTPT-----PTPTPTGOTPTTPTTPTT----- 3654
Qy 1594 RKGIREVYFVYNAKPAIDIMYPSPRPFTGITWRYRLQTVKSLAGVSLMLRLMASLR 1653
Db 3655 -----VPTPTPTGOT-----QPTTPTPT----- 3673
Qy 1654 WDDMAKVPFGGSGSTFTETSEITTEIIRKRDVGPYIGREYCIKIIICPIGV--E 1710
Db 3674 --TTTTVPTPTPTGOTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 3720
Qy 1711 TPKEPTPTQKRLRSALR-----KRPEPKQGT-----PVIEETWVAEELMEIR 1758
Db 3721 TTTTPTPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 3771
Qy 1759 RAFAERVEKEKAQAVEQAKRLEQOKPTVIATSTTSPRS--STTSTISP-----AQK 1809
```



```

Db 709 -----LQDEKOVANDKIGKIEIKAIKDIDAATTNQAVEAIKT-----KAIINDIQTTP 756
Qy 1130 ISRAMDFESKLCQDSESNSTLNSSDPTYSIODSSEEDMIVONSNESISQOFPTREODEVY 1189
Db 757 ATTA-----KAAALEEFDEVVQAQ-----IDQAPLNDP-----TTNEVEAIA 793
Qy 1190 LEPPLCELVSQ-----ESTGNCEDRLPVKCTE-----ANGKRPQOQKLEERPKNKCDQIK 1241
Db 794 IERINAAKVSQKAIETATTAQDLERKNEEISKIENTITDSTQTKDAYNEVQAATARK 853
Qy 1242 LKNTDDKNNRESEKQORTSTFOINKDNKPKIYKCECLKEISESEVAVSGNEPKV 1301
Db 854 AQMAT-VSATATEVEVAEADAADAQKQCHDIQVYKSK-----QEVADIK-----SKILDKI 905
Qy 1302 NNIN---KIIPENDIKSLTVKESAIRPFLNGVIMEDFERNSSSETKSHLSSDAEGNV 1358
Db 906 NAIQOAKVPAAD-----TEVENA-----YNTKQEOIONSASSTEEKQAAV 948
Qy 1359 RSLEFLPSTRKESDSTQ-----TTTSPASCPESNS---VNOVEDMEIETSEVKVYSSPTT 1411
Db 949 TE-----LDTKKQOARTNLDAAANTNSDVTTAKDNSIAIMQVOAATTKKSDANA----- 997
Qy 1412 SEESNLNDPIDENGLPIKNNENNVNGESKRKTVITEVTMTSTVATESKTVIKVEKDK 1471
Db 998 -----EIAOKASEKTAIEAMNDST-----TEEQOAK-DKVDQ 1030
Qy 1472 QTVVSTTE-----NCAKSTVTTTTVTKLSTPSTGSDVDIISVKEOSKTVITVTD 1524
Db 1031 AVTANADIDNNAANDVDNAKTMTNATIAAI-TPDA-----NVKPAKQAIAKVOA 1082
Qy 1525 SLTTGGTLVTSMYKSKSTPDKYKLMKFSRPKTRSGTAL-PSYKRVYTSKTSSTIV 1583
Db 1083 QETAIDG-----NNGSTTEEKAQAQOQVTEKTADAIDAHAETNAEVEAKKA--- 1131
Qy 1584 LPNDLKLARKGGIREVYFYNNAKPAIDIMYPSRPFTGTTMYRILQYVKSIAQVSL 1643
Db 1132 -----AIAKIEAIOPATTTQDNKEAI----- 1153
Qy 1644 MLRLMASLRWMDMAKVPGGSGSTRETSETETTTETIKRBDVPGYCIREFYCIKII 1703
Db 1154 -----ATRKANERKTAIAQOTDITAEITAANAADVMATQAN 1190
Qy 1704 CFIQVPEYKKEPTPORKGLRSSALRPKRPETPKOTGPVLIETWAAEEL-----ELME 1757
Db 1191 SNIEANSQNDVQAKTGENSI-----DQVTPYVKNKATARNKITAITAILNKKLOE 1240
Qy 1758 IAPAEVYKEKAQAVEOQAKRLEOQKPTVIATSTSTSSSTSIISA-----QKVVV 1812
Db 1241 IQATPDATDEEK-QAADAENANTENKANOAISATTTNAQVDEKAKAAEAIAINAVPEKVVK 1299
Qy 1813 ABISGSVTTGTMMVLTTKVGSPATVTFQONKNFHQTFATWVKOGSN-----SGVVQV 1865
Db 1300 KQAADEIDQATQOTNVINNDQNAATTEKEKALIQOLATAVDAKNNITIAATDDNVDQA 1359
Qy 1866 QOKVIGIISSTGTSQOTFTSPQPTAIVTIRPNTSGSGTTSNSOV---ITGPQIRPKMT 1923
Db 1360 KD-----AGKNSIOSTQTPATVAVSNKNDVDAQVTTQNAIDMTTG----- 1400
Qy 1924 VIRTPLQOSTLKGAIIIRPVWVOPGAPQOVW-----TOIIRQOPVS-TAVSAPNTV 1973
Db 1401 ---ATTEENNAKDLV---LKAKEKAYODILNAQTNDVTO-IKDDAVADIGITADDTT 1453
Qy 1974 SSTPOKSLTSASTSNIOSSASOPPRPOQOVKLTLM-AQLTOLTOGHGNGGLVTVIG 2032
Db 1454 KQVAKDELATKANEKALIAQTADATTEKEKANOQVADQLOGNONINMAOSIDVWNA 1513
Qy 2033 QGOTTGQOLIPQGVTVLPQGOQLMQAAMPNGVQORFLFTPLATTAT----- 2081
Db 1514 KQNAIQADIPQIASTVDKTNARAEILT-----EMQNKITEIILNNETNEKNDIGPV 1567
Qy 2082 -----ASTTTTSTTAAGTQGRQSKLSQOVQVODKTLPPAOSSSVCPANA 2129
Db 1568 RAYEBGLNNINAAITTTGDV-TTAAKDTAVQK-----VOOLHANPVKKPAKRKEIDQAA 1620

```

```

Qy 2130 QPQTQAPSAPOQOTOPQSPAQPEVOTQOTVSSHVSEADPTHAQSSKPPVAAQ 2189
Db 1621 DKKT-QIEQTPNASSQOEDINDAKQOEDVTELN-QAKTVWDOSSTNEVDNAVKEGAKINA- 1677
Qy 2190 SQOPNSVOGQSPVRYQSPQTRI-----RPSPSQLSPGQOQV-----QTTT 2232
Db 1678 VKFSEYKKDALAKIEDAYNAKVENADNSNASTSSIEAKQKLAELKOTADONNVQATS 1737
Qy 2233 SOPPIQPIPTSL-QIPSQOQSPQPOVQOSTQOTLSSGOTLNOVYSSPSRPOLOQOPQ 2291
Db 1738 KDDIEVQIHNDDNINDYITIPGKKESATTDLYAVADQKKNISADPTNTQ----- 1788
Qy 2292 QYIAPVLOQOQOVVLSQIOSQVVAQIQAOQSGVPQOIKQLPIQIOQSSAVQTHQIONV 2351
Db 1789 -----DEKQOAIKQOVQUNQVQTALESINGV-----DNGDVVDALQGRK 1827
Qy 2352 TVQASVQSOLOQVQOLRQOQKQKQOQEIKEHRTLOASNOSEIIQOVYKNNVIEH 2411
Db 1828 AIDAIOVDATV-----KPKANOALEVKAEDTKESIDQSDQLTAEKTEALAMIKO 1877
Qy 2412 L--KOKSMTPA-----EREENQRMIVCQVAKYILDKID----- 2444
Db 1878 IYDQAKQGITDATTAEVKAQAQGLEAFDNIQIDSTERQKAI---EELETALDQIEAGV 1934
Qy 2445 -----KEKQA-----AKKRKRESEVEDKRSKONAT-KISALLFHKHQLRAEIL- 2488
Db 1935 NVNADATTEKEKAFNALIEDILSKATEDISDQTTNAIATVKNKSAI-----BOLKQORIN 1989
Qy 2489 ---KRAL-----LDKDOIQIEVQELKRDKIK-----KENDMQLQATNAV 2527
Db 1990 PEYKKNALAIREFVVK-QIEIIRKADADASAKELIARDLGRFYRFDKIDKTQTNAE 2047
Qy 2528 AAPCPVPTVLPAPAPPSPPPGVOHTGLSTPTLPVASQKRRKEEKSSSKKK 2587
Db 2048 VAELOWVT--IPAIEIIVQNDPDANDTNNGI-----DNDNATANSNAN 2089
Qy 2588 KMISTSKETKKDT 2601
Db 2090 APTWENGQPNVSET 2103

RESULT 12
US-08-681-219-32
; Sequence 32, Application US/08681219
; Patent No. US20020058607A1
; GENERAL INFORMATION:
; APPLICANT: Takaaki Sato and Junn Yanagisawa
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GLGF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/681,219
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JKM
; TELECOMMUNICATION INFORMATION:

```


TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-681-219-32

Query Match 2.3%; Score 336; DB 8; Length 2843;
Best Local Similarity 17.1%; Pred. No. 3e-07;

Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;

QY 387 EIRHMDITEDLTNKGAGSKSFLAA-ANEELIESIRAKKGIDIDNKGSPETEEDKDKN---442
DB 802 DTRHRDNRSDNNT---GMMTVLSPLNTTVLPSSSSSGSLDSRS---EKDNLER 854
QY 443 -----ETENDSKDAEKNREFEEDS-----LEKSDDKTPDDPEQCK-----480
DB 855 ERGIGLGNYPATENPCTSKRGLOISTTAQAQAKVMEVSAIHTSQEDRSSTTELHC 914
QY 481 -----SEVGDf-KSEKSN-----GELSESPGAG 502
DB 915 VTDERNALRRSSAAHTHSNTNTNFKSENSRNCMPYAKLEYKRSSNDLSNVSSSDGYG 974
QY 503 KGASGSTRITRLRNPDSKLSQLKS--QQVAAAHAEANKL-FKEGKEVLVNV-----551
DB 975 KRQGMPSIESYSEDESKSCSYGYRADLAHSHSNHMDNDGELDPIPNLSLKYSDE 1034
QY 552 --SOGELSTRSTKEVIMKGNINNYFKLQEGKYRVYHNOYSTNSPALNKHQREDHK- 608
DB 1035 QLSNGQSPSONRMARPKHIEDETIKQSEQRQR--NOSTT--YPVYESTDDKHLKF 1089
QY 609 RRLLAHKFCILTAGEFKKWSVHGSVLTSTLRITITQLENNIPSPFLHPNMASTRAM 668
DB 1090 QPHFGQCEVSPYRSRGANS-----ETNRVGS-----NGLNQ 1123
QY 669 IKAVMQSKPREFALALALECAVKEPVMLPIWREFLGHTRLHRMTSIEEREKVKYKKE 728
DB 1124 NVQSGLQED-----DYEDDKPNTYSE 1145
QY 729 KQOEETMQOATWVKYTPVKHQVKQGEYRVTYGQSWISKTHYRVYKPLPGNT 788
DB 1146 KXSEEE-----QHEEERPTNY 1162
QY 789 NVNYRKSLEGTCKNMNDENDES-----DKRCKSPKKIKITIEDPSEKD 831
DB 1163 STAYNEE---KRHVDPIDYSLIKATDIPSSOKQSPFSKSSSGSKTEHMSSEENT 1218
QY 832 EYKSDAAGADQNDMDISKITEKKDQDVKEELDSDS-----DKP-CKEPPA 877
DB 1219 STPSSNAKR--QNOHLPSAQSRSQGPQKAATCKVYSINQETIQYCYVDPIICFSRCS 1275
QY 878 EYVDDMKTESHVNCQSSQ-----VDVYVNVSEGFHLRT-----SYKKTKTSS 919
DB 1276 STLSLSAEDETIGCNOTTQEADSANLQIAIEIKETIGTSAEDPVEVPAVSQHPRTKSS 1335
QY 920 KIDGLLEIRIKQFTLEKORLEKIKLEGGIKGIGCTJSTMSKN-----LSESVI--TKAK 973
DB 1336 RLOG-----SLSSESAAHKAVERFSSGKSPKSGAQTPKSPHYVQEPPLMPSRCT 1388
QY 974 EGCQSDMRQEOSPANNNDQPEDLIQGCQSQSDSVLRMSPSHHTTKILYPKDVLVDVSI 1033
DB 1389 SVASLDSFESRSIASSVQSP-----CGSMVGITISPLDLPSPQOTMPPSR-----1435
QY 1034 RSEPTGCPKONSTENDIEEKVSDLASRGQPTKSKTGKDNDFETIDSKLASAD-----1085
DB 1436 --SKTTPPPPTAQTAKRE-----VPKNKAPTAEKRESC-----PKQAAVNAAYORVOY 1481
QY 1086 --DIGFLI-----CKNK-----KPLQDESDTIV-----SSKSALHSSV 1118

DB 1482 LPDADTLHPATESTPDGFCSSSLALSIDPEFIQKVELRIIMPVQENDNGNETSEQ 1541
QY 1119 PKSTNDRDATPLBRADPECKLGCDSFNSNTLNSSDPVYSIODSPEDMIVQNSNISIE 1178
DB 1542 PKSENMQ-----EKEAEKTTIDSEKD--LLDSDDD-----1570
QY 1179 QFRTQDEQVEVLEPLKCELVGSGSTGNCEDRLPVKGTTEANGKAKPSOQKLEBRPVKCS 1238
DB 1571 -----DIEILE--EC-IISA-----MPTKSR--KAKKPAQDTASKLPPVARKPS 1610
QY 1239 QT---KIKNTDKKNNEBRSEKKGQRTSTFQINGKDNKPKYILKECELKEISESVYSG 1295
DB 1611 QLPVYKILPSONRLOQKHVSFTPG-----DQMPRYV---CV-----EGYPINE 1651
QY 1296 NVBPKVNNIKIIPENDIKSLTWKESAIRPFINGDIYMEPFNRSSEFKSHLLSSDAE 1355
DB 1652 STATSLSDLTIESPPNL-----AAGGVAGAGQSGEFERKDTIPIEBGR--STDEAQ 1701
QY 1356 GNYRDSLETLPSYTES-----DSTQTTTPSASCP 1384
DB 1702 GGRKTSYV-TIPELDNNKAEBGDILAEICINSAMPKSGSHKPFVKKIMDQVQASASSAP 1760
QY 1385 ESNVAVQVEDMELETSEVKKVTs-----SPTSEBSNLN 1420
DB 1761 NKNQLOGKK--KKPTSPVKIPQNTETRYTRVRKNADSKNNLNAERYVSDNRKQKONLAN 1818
QY 1421 DFIDENGLPIKNKNENYNGE-----SKRR 1443
DB 1819 NSKDFNDKLPNNDRRGSAFDSPPHHYPIEPTPCFNSNDLSLDPDDDDVDLSREK 1878
QY 1444 TVITEVYTMSTVATVATESKTIVKVEKDKQTVASTENCASVTY-----TTTTTV 1493
DB 1879 ---AEIRAKAKENSEBAKVTSHELTNSOOSANKQOIAKQPINRQCPPILOKQSTFPQ 1935
QY 1494 TKLSTPSTGSDVIDIIVKEDSKTVVTTYDLSLTGGTGLVYTMYSKEVSTEDKVKLAK 1553
DB 1936 SSKDIDPRGAAD--EKQNFALIENTPVCFs--HNSSLSDSLDQENNNKENEPIKE 1989
QY 1554 FSRP-----KKTSGTALPSYRK-----EYTKSTKSIETLPNDL-----1589
DB 1990 TEPPDSQGEPSKPOAGVAPAKSFHVEDTPYCFERNSSLSLSDSDLLQECISSAMPK 2049
QY 1590 -KTLAKGCIREVYVFNNAKP-----ALDIWPPSPRPFGIT-----WRY 1630
DB 2050 KKKPSRLKGDN-----KHSPPRMGILGEDTLTLKLDIQRPDSHGSLSPSENDWKA 2103
QY 1631 RIQTVKSLAGVSLMLRLMAFLWMDMAKVPYGGSTRTFSEETITTEILIKRDVAP 1690
DB 2104 IQEGANSIV-----SLEHQAAAACL-----SKQASDSD--STLSLKS---2140
QY 1691 YGIRFEYCIRKILCPIGVETPEKETPTPQRKGLRSALRPKRREPTEKOTGPVLIETVAAE 1750
DB 2141 -GI-----SLGSPHLHPDQEKPTFSKGR--ILNPGKEST-----LETKKIE 2182
QY 1751 EF-----LELWIRAFAREVEKEKAQAVEQAKKRLQOKP-----TVIAT 1791
DB 2183 SESEKGIKGGKVKYKSLITGCVRSNSE-----ISGQMKQPLQANMPSISGRRTMIHI 2233
QY 1792 STTSPSTSTSPQAKWVAPISGVYTGTFMVLITTVGSPATVTPQNNKHFQTFAT 1851
DB 2234 PGVRNSSSTSPVSKGPPKTPASKSPSEGQATTSPPRGAKPSV-----2278
QY 1852 WYKQGSNSGVVQVQKVLGILIPSSYTGTSQQTFTSFQPTAVVTLRPNTSGSGGTYSNQ 1911
DB 2279 -----KSELSPVARKTSQIG-----GSSK-----APSRGSGSDSTPS--2310
QY 1912 VITGPQIRGMVYIRFPLOQSTLGKAIIRTPVWVQPG---APQVMTQIIRGQPVSTAV 1967
DB 2311 -----RPAQOPLSRPIOSPGRNS-----ISPGRNGISPPNKISOLPRTSSPSTA-2354
QY 1968 SAPNTVSTTPGQSKILTS-----ATSTSNIQSSASQPPRPQOQVQLTMAQLTQLQ 2018
DB 2355 ---STKSSGSGKWSYTSPPGRMSQONLTKQTGLSKMASSIPSEAS---KGLNOMNN 2406

Qy	2019	GHNMOLTVIIGQGGOTQZQL-----LIPQSVTVLPGGQQLMAAPNGVQZFL	2071
Db	2407	GGNANKKVELSRMSSTSSSESDRSEPRVLAQSTFIKAQSPTLRRKLEESAS----	2461
Qy	2072	FTPLATATTAATTTTATTTTAAAGTAGEOROSKLP-----QKQVHOD-----KLLPBAQ	2120
Db	2462	FESLSPSSPRASPTBSQAQCPV-----LSPSLPMSLSTSHSVQAGWRKLLPNL	2511
Qy	2121	SSSYGPKAKPQIAQPSARQPOTQPSAPQAEVQIQPQEVQIQTTTSSHVSEAPQTHAQ	2180
Db	2512	SPTEFYMDGPRAKRHDIARSHSESPRLINRGTKRE-----HSHSSSLPVSFTWRRTG	2568
Qy	2181	SSKPOVAQSQPQSNVQGGSPVSVQSPQTRIPRSPSOLS--PGQSOQVQTTSSQPIPIQ	2239
Db	2569	SSSSILASSESSBKAKSEDEKHNYSISGT--KQSENVYSAKGITWRIKIKENFSP-----	2622
Qy	2240	PHTSLQIPSGQPOSQPOVQSSQTLLSGQT	2270
Db	2623	-----TNTSQTVSSGAT	2635

```

RESULT 13
US-09-902-432-4
: Sequence 4, Application US/0902432
: Patent No. US2002016002a1
: GENERAL INFORMATION:
: APPLICANT: Irwin H. Gelman
: APPLICANT: Susan G. Jaken
: TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
: FILE REFERENCE: A30558-A-EWC-A 070156,0597
: CURRENT APPLICATION NUMBER: US/09/902,432
: CURRENT FILING DATE: 2002-04-08
: PRIOR APPLICATION NUMBER: 08/978,277
: PRIOR FILING DATE: 1997-11-25
: PRIOR APPLICATION NUMBER: 08/665,401
: PRIOR FILING DATE: 1996-06-18
: PRIOR APPLICATION NUMBER: 08/635,121
: PRIOR FILING DATE: 1996-04-19
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 1596
: TYPE: PRT
: ORGANISM: Rattus norvegicus
US-09-902-432-4

```

Query Match 2.3%; Score 323; DB 9; Length 1596;
Best Local Similarity 18.6%; Pred. No. 5.9e-07;
Matches 374; Conservative 261; Mismatches 671; Indels 704; Gaps 87

```

QY 110 1ERKBAKVAANANCOELEEETMOOAIWVITFEPAKHQVAKOAGEEYKRVIGGWSMISKT 115
Db 89 VGOESESDEYREKDR -VEEMANSTAVE-----DIYKQOE----- 123
QY 776 HWYFEVLEGNFNVMYRKLEBTKNMNDENDESDRKSCRSPKTI-----KIEPD 827
Db 124 -TSEIIQID-----ASENNVEEMVQPAESQANDYGEKFKVFEKFEKTVAKD 170
QY 828 S-----EKDEYKGSDAKAKADDN-----EMDISITKEKKDQV 860
Db 171 KNEKSDTVOLLTVAKKKBGEBAEASVAGAHQBPSEYATAGBASAKESBELKOSTEKQEBTL 230
QY 861 KE-----LLDSDSPCKPEEPMPEVDDDMKTESHNCOE--SQADVAVNASE----- 904
Db 231 KOEOSTEIRILOAESQOAEAEAKDGECEKQKEPTKPSPESSPVNSETTSFKKFFTH 290
QY 905 ---GEHLRTSYKKKKTSSKIDGL--LEBRIKQFTLEBNORLEKILBEGTIGIGKTSTNS 959
Db 291 GMAQWRKKTISFK-----SKEDDLTAEKR-----KEDEAKKVDE----- 326
QY 960 SKNLSSEPVITAKCEGQCSQSDMRQESPPANNDOPEDLLQGCQSQSDSSVLRMSDPSHTTN 1019

```

Db	327	----	EKKTEPASEQEOPADDTQ	-----	ARLSAD	-----	352
Qy	1020	KLYPEKRVLDVDSIRSEPTKCPKONSTIENDIEEKVSDLASRQOEPTKSKTGNDFIDS	1079				
Db	353	-YEK-----	VELPLEQVGDLEASSEE	-	KCAPLATEVFN	-	DE 385
Qy	1080	KLASADIGITLCKNKKPLIOESDITVSSSKALHSSPKSTNDRDATPLSRANDPECK	1139				
Db	386	KMEHQEV-----	VAEYVSTVEKTEEEOG	-----			411
Qy	1140	LGCSEMSSTJENSJDVTIODSSEEDMIYQNESISISOPFRPRODQVLEPLKCELVS	1199				
Db	412	-GGEHGGVVEGTGESLPPREKLAPEQVPOEA	EPAEELMSKREM	-----	CVS	458	
Qy	1200	GESTGNCEDRLPVYGTAEANGKKPSQOKKLEERPVNKCSDOIKLKTNTDKNNENRESEK	1259				
Db	459	GGDHQTJLTDLSP	-----	EKKTLPKHPEGIVS	-----		484
Qy	1260	GORTSTFOJNGKNKPKYIYKGECLKEISRYVSGVNEPKNNINKIIPENDIKSLTVK	1319				
Db	485	-----	EVEMLSQERIKKVOGSPLK	-----	KLFSSSGLKLIKSG	517	
Qy	1320	ESAIRPFIINGIWMEDNENNSSEPTSHLLSSDADAGNRSOLETLPSTKESDTSQTTTP	1379				
Db	518	KQKGRGGGD	-----	EEEGEYOHHTES	-----	PSADQKCES	552
Qy	1380	SASCPESNSVQVDEMEIETSEVKKYTSPITSEESNLSNDFIDENGLPJIKNENVAGE	1439				
Db	553	SASSP	-----	EEPEPTTCLKGLPEARQODEAEGETTSD	-----	GE	588
Qy	1440	SKRRTVITEVTMTS	-----	TVATESKVIYKVEKGDQKQVTSSTENCAKSTVTTTTVTNLS	1497		
Db	589	KKREGI	-----	TPMASFKMVTPKPKRVRRPESDKE	-----	ELEKVKASATLSDTSVSEMO	641
Qy	1498	TPSTGSGVDIISVEQOK	-----	TYVTTTVYDLSLTGGLTVMYTSKESFTSDAKKL	1551		
Db	642	D-----	EYKTVVEEKKPEPKRRDVTJSWEALLCVG	-----	SSKKRRKASSDD	687	
Qy	1552	MKFSPPKTRSGTALPYRKFVTKSTYKTSIFVLPNDLKKLARKGIRFVPFNNAKPA	1611				
Db	688	---EGEPRTLAG	---DSHRADEASKDEA	-----	GTDAPVASTQEDQQA	725	
Qy	1612	L-DIMPYSPRPTEGITMYRRLQTVKSLAGVSLMLRLMASLRMDMAKYPGGGS	1667				
Db	726	QGSSEPAGSPSEG	-----	EGVS	-----	TWSEFKRLVTPRKKSASK	762
Qy	1668	---TRETSEIETTTELKRDYGPYGIREFYIKRIICPIGVPEPKETPPPOAKGLRS	1725				
Db	763	LEKRAEDSSVEOLSTELTEPSRE	-----	ESWVSIKKFI	-----	PGRRKKRA	802
Qy	1726	SALRPKREPPIKONGPVIETVWAELELELMEIRAE	-----	RYEKKQAQAVEQAKKRLQ	1783		
Db	803	DG---KQOEAIVEDGPEIN	---EDDPNPAVAPLSELTNAERER	---MEAOGNTLQ	853		
Qy	1784	OKPVIATISTSPSTSTSTISPAQVWMAPIGSGVTGTKAVLTLTKVGPATVFOQK	1843				
Db	854	LLGAVVYSELS	-----	KTIVHTYSVAANIDTRAVTSEBSPSWISASYTE	900		
Qy	1844	NFHOTFATWKGQSGNSGVVOOKVLGIIIPSTG	-----	TSOQFTS	1886		
Db	901	PLEHT	-----	AGEAMPVEVEYTER	---DLIAETPTVLNLTLEPKGDADHDMDTSEVDTFS	952	
Qy	1887	FQPTATVYTR	-----	PMTSGSGTNSOYITNGPOIRPQMYLIRPLDQSTLGRKII	1939		
Db	953	EAVTAETSEALTEEVTASGAELTTDMSAVSOLTDSPJTTEAPVPEVEBEG	1007				
Qy	1940	RTPVMPVOPGAPOQWMTIIRGQVSTAVSAPNTVST	-----	PGOCSLTASTST	---SNI	1991	
Db	1008	-----	VLDTBEERQYQAIL	---QAVNAKVAEESQVPAITQYQVQRTSGKALEKVEVEBSEV	1061		
Qy	1992	QSSASQPPRPOQGVKLTMAQLTQLTQHGNGMGLTVIYQGGQGTGGLQLIPQGVTVLP	2051				
Db	1062	LASKEKDVMKGVQDEAGAE	---HLAOG	-----	SETGQATPSSLSEPEVTAAYD	1108	


```

Db 1037 LMOODPKSVRL-----PVSQ 1052
QY 2652 STEDAMTVLTPTLTK---DYEGLKRVLSLOAHKMAWPLEPVDNDAPDYGYIKEPMD 2708
      : | | : : : | : | | : | | : | |
Db 1053 NVHPPRGFLNPDSDQRMPMQSGSVPMVSLQG-----PASVPPSPDKO---RMPMP 1100
      : | | : : : | : | | : | | : | |
QY 2709 LAT-MEERVOQRYYEKLTEFVADMTKIFDNCRYYNPSDSPFYQCAEVLES 2757
      : | : : : : : | | | : | : :
Db 1101 VNTPLGNSNRKMYQESPO-----NPSSSPLAEMASLPEA 1135

```

Search completed: November 20, 2002, 16:45:59
 Job time : 88.712 secs

US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 3.0%; Score 432.5; DB 4; Length 2137;
Best Local Similarity 17.2%; Pred. No. 3e-17;
Matches 395; Conservative 355; Mismatches 999; Indels 547; Gaps 77;

QY 394 ITEDLTNKAAGSNKSFIAANEI-----LESTRAKKGIDIDNKSPEETEKD- 440
DB 149 MSSSVTLTDDSEKAGISQSSSETSNOSSELTNYASTDHESTTNDNTAQQQONKSSN 208
QY 441 --KNETENDSKDAKNEEFEDQSLKSDDKPTDDPEQKSEVQFKSEKSNGLSES 498
DB 209 VTSKSTQSNMTSSSEKSNISNLQSIETKATDSLATSEAKTSTNOISNLTSTSNQSSPT 268
QY 499 PGAKGAGSGSRITRLRN-----DSKLSQKSOQVAAAAHEAKLF-----KEG 544
DB 269 SFAMKRTFSRPTVNTMAAPTSTTTTSLTNSVYVKNKNFNEHMLSGSATVDPKGTG 328
QY 545 KEV-----VNSQGEIS---RLSTKEVIMKGNIN---NYF----- 575
DB 329 TATLTTPAISKGAISLNTRLDSNRSFRFTGKVLGNRYBESPDGVTGGDGTGAFATSPG 388
QY 576 --RLQGEG-----KYRVYHNOYSTNSFALNKHQRE-----D 605
DB 389 PLQIGKEGAAVIGIGINNAPFKLDITYHNTSTPKSDAKADPRNVCAGGAFVSTSD 448
QY 606 HD-----KRRHLAHPCLTPAGE-----FKNGSVHGSKVLTITLRLITIQLENNIPSS 655
DB 449 RGNMATTSSAAKLNVQPTDNSFQDFVIDYNGD--TKVMTVTYAGQTFTR----- 497
QY 656 FLHFNMAASHRANWIKAVOMCSKPREPALALAILBCAVPVYMLPIWREFLGHTRLHMTS 715
DB 498 -----NLTDWIKN-----SGTTFISLMTASTGAKNLQOYVQCFTEYTESAVAKVRY 545
QY 716 IEREK-KEKYKKKEKKKEEETMOATVWVYKTPVKHOV--WKQKGEERYVT-----GY 766
DB 546 VDATMGKDIIPKTIAGEVDATV-----IDKOLNMLKNSGYSYVSTDALQNSNY 595
QY 767 GGMWMI-----SKTHVYRFVFKLPQNTNVNRYKSLSEGTKNMDEMMDMSDKRKCRS 818
DB 596 SETSGTPTLKITNSSQVITYKF-----KDVQGPQISVD-----SQTREVGKT 637
QY 819 PKKIKI-EPDSEKD-----EVKGSDAKAGADQ-----NEMDISKITEKKDDVEXLL 864
DB 638 INPITITTTTDSKDVLTITTVGLPSGLSFDOTWTITIGTSEVGTITVY-----VNT 690
QY 865 DSDSDKPKKEPMEVDDMKTESHVNCOESSOV-----DVYVNS 903
DB 691 DATGNVTSKQPTITIDTISPVNVVTPSOASEVFTPIPIITITADNSGKVVHTVIGLPL 750
QY 904 EGFHLRTSYKKKTKSSKLDGLLEKRIKQFTLEKQORLEKIKLEGGICIGKTSNNSNLT 963
DB 751 QGLKFDASTNSIVGTPQIQGTNTTITESTDASGNKTYTKRIYE-----VTRNSASDS 802

QY 964 SESPVITKAKEGQ-----SDSMROEOPSNANNDDPEDILJQCGSQSDSVYLKMSDPESHYT 1018
DB 803 TSTSVNSVSTSNISNLSLSDSVKASQSLSTSKLSLSL--SASTSNSTIQASSEASTS 860
QY 1019 NKLYPKRDVLDVYSIRSPETKCPKQNSIENDIEEKVSDLASRQOEPTKSKTKGNDFFID 1078
DB 861 KQL-----SEFASSTSDSASARKSESTSKTS-----LSE 893
QY 1079 SKLASADDITLICKKPKPLQIESDPTIVSSKSAHLSVPKSTNDPDAIPLSRAMDFEG 1138
DB 894 STSTSVSDASASVSTSES---ASTSTVSGSTSTSIDSTSTSDASATKASEAS-7S 948
QY 1139 KLGCDSESNSTLENSDPTVSIDS-SEEDMIVONSNEISEOPRTREQDVEVLEPLKCEL 1197
DB 949 KLISESVSTSDSASTSVSDSNKASTSLSKSTSVSDSTSTSD----- 997
QY 1198 VSGESTNCEDRLPVKGTANGKRRKPSQAKLEERPVKKCSDOIKLKNTTKNNENRESE 1257
DB 998 --SASTSTSE-----SESDSASTSLSESTSTSVSDSTSTSTSDSASASASESE 1043
QY 1258 KKGORTSTPQINGKDNKPKIYLKGECEKISEKRVSGNVEPKVNNINKIIPENDIKSLT 1317
DB 1044 SNSKSTIS-----LSE-STSLSGSTVAST-----SDSASTS 1074
QY 1318 VKESAIRPPIINGDIYIMEDFERNSESTKSHLLSSDAEGYRDSLETLPSTKESDSTQTT 1377
DB 1075 TSESE-----SDSTSTSLSESTSTSLSGS--TSASDSAST--STSESDSTSES 1120
QY 1378 TPSASCPESNVNOVEDMEITSEVKRYVTASPLTISEESN-----LSNDFIDENGL 1428
DB 1121 T--SLSESLSTSVSDSTASSTSE---SASTSEESNSASTSLSGSLSTSIDSTST 1173
QY 1429 PINKENYVN--GESKRYKIVTEVTMTSTVATESKTVIKVEKGQKQVWVSTENCASSTVT 1487
DB 1174 STSDSASTSTSESESDSTSTSLSTSTSLSDSTSTSTSEASASTSESDSTSTSLSE 1233
QY 1488 TTTTTVTKLSTPSTGSGVDI--ISVKEQKTVV-----TTVTVDLSLTGTGGLTSMTVS 1540
DB 1234 STSTSVSDSTSVSDSASTSVSDSASTSISESLSTSVSDS--TSTSTSDSASTS 1292
QY 1341 KEYSTROKVKIMKSRKRTKTSGLALPSYKRFVYKSKTSIFVLJPNDDIKLAKKAGSIRE 1600
DB 1293 ESDSTSESTSLSEISSTSVSDSTSA--STSDSASTSLSES-----ESDSASTSLSG--- 1341
QY 1601 VPYFNNAKPALDIMPYSPRPTGITWRYRLQYVKSSTLAGVSLMLRLIMASLFRMDMAAK 1660
DB 1342 -----STSTSLS----- 1348
QY 1661 VPPGGSTRETSETTEITTEILIKRDVGPYGIREFYCIRKIICPIGVPEPKETPTPQR 1720
DB 1349 -----DSTSTSDSASTS-----ESDSERASTSL 1375
QY 1721 KGLASSALRPKRPEPTQGTGV-----IIFETVAEELELMELRAFAER 1764
DB 1376 SGTSTSTSLSDSTSTSDSASTSVSDSNASTSLSGSLSTSVSDSTSTSDSAST 1435
QY 1765 VEKKAQAVEQAKRRLKLEOKRPVIATSTPTSTSTSTSPAKVWVAPISGVTGTGK 1824
DB 1436 SEDS-----SBRASTSLSGSTSTSIDSTSTSDASTSVSE-----SNSTSTSIS 1484
QY 1825 MVLTKVSGPATVTFPOONKNEHQTFAIWKOGOSNGVVOQAVLGIIBSSTSGQTF 1884
DB 1485 ESLSTSV--SDSTST-----STSDSASTSVSDSASTSSSESV-----STSDSEST 1532
QY 1885 TSPQPRATVYI-----RPNISGSGGTTNSNQVITGQIRGGMVITPPLQOSTLGRAII 1939
DB 1533 TSTSDASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSDSAS--ASTSESDSAST 1589
QY 1940 RTPVWQPGAPQOQVMTQILIRGQPVSTAVAPNTVSPGQK---SLTSASTSNIOSSAS 1996
DB 1590 SSSSESVSTSVSDSTSVASSTSEASTSTSVSDSNASTSLSESTSTSLSDSTSMSTSDAST 1649
QY 1997 QPPRQOGQVKLTMAOLTQLTQGHGNGGLTVVIQGGQGTGGQLQLIPQGVTVLPGFGQ 2056


```

QY 1460 SKTYIKVE-----KGDQVYSSPENCASQVTTT-----TTVTKLSTPSTG 1503
Db 1615 SNGHEHLELMPDAHKSDAKOSIDNKYNDOSNTINTTTPATBEEKOKALDKLAKDAG 1674
QY 1504 --SVDIISVEQ--SKTYVTTTVDLSLTGTLVTSMTVSKREYSTRDKVL-MKFSRP 1557
Db 1675 YNKVDAQOTNOQVSDAKTEALDITTN-----IQANVAKKPSAR--VELDSKPEDL 1722
QY 1558 KTRTSGLALSRYKRFVTKSKKSLFVLPN--DDLKTLARKG-----GIREVP 1602
Db 1723 K--ROIWATPN--ATEEKKODAIORLNCKRDEVKMLINODRRDNVEHCKHNGIOELE 1776
QY 1603 YFNNAKPAIDIMPYPSRPFGITWRYRLQTVKSLAGVSLMLRLWASLRMDMAKVP 1662
Db 1777 TIHAN-----PFRKSDALOELQOT-KFISOTELI-----1803
QY 1663 PGGSSTPSTSETEITTEIIRKRDVPGYGIREFYCIKRIICPIGVETREKPTPQKRG 1722
Db 1804 ---NNKDATNEEKDEAKRL-----EISKKTITTNINQATNNQVNDANKDG 1848
QY 1723 LRSSALPKRPETPKQGPVLIETWAEELMEIRAFERVEKEKAQAEQAKRLE 1782
Db 1849 MNEIA--TIPATIKTDKATAIDKKAEOVTI--INGNDATDEEKAERKLVEKAKIE 1904
QY 1783 Q-----OKPVIATSTPSTSTSTISPAQVAV--APISG 1817
Db 1905 AKSNITSDTREVNAKAKTNLEKINNIQSTQTKTNAKEINKKAEOLQIINTP--1961
QY 1818 SVTTGKMWLTTKVSPATVTEQONKNPHOTFATWVKOGSNGVVOOKVGLIIPSSST 1877
Db 1962 DATEEKKOEAIRNAGIQAQIQUINNASHY--OEVNESKTN--IATIKSVOPVNIKRP 2018
QY 1878 GTSQOTSPRPRT-----AT-----VT-----IRPNTSGSGTTSNSQ 1911
Db 2019 AINSULTQANOKTLLIGNDNATDEKEAKQVLTQKLINEQIHKSTODNOVDNKAQ 2078
QY 1912 VITGPQI-----RPGMTVIRTPLOOSTLTKALIRTPVWVQCAPQVMTQIIRGQV- 1963
Db 2079 AITAIKLINAMAHKRQDAINILTLMLASK--KSDIRAN--QDATTEKKNIAI--QSID 2130
QY 1964 STAVSAPTVSTPGOKSLTSATSTSNQSSASQPPR-----PQGGVKLTMAQLTQLTQ 2018
Db 2131 DTLQARNNING-----ANTMALVDENLEDKQRLQRIVLSTQRTQAKADIAQAI--2181
QY 2019 GHGNGOGLTVIOGOGTGTGLOLIPQGVTVLPGRGOOLMOAMPNGVGFLLTPLATY 2078
Db 2182 ----GQQRSTIDONATTEKQKALELNOETNGVNDRIQALALANQVNTBERKNILETI 2237
QY 2079 ATT-----ASTTTTVSTTAAGTGEOROSKLSPOQVHODKTLPPAOS 2121
Db 2238 RNVEPIYIVKPKANEIIRKKAEOQTTLLINQODATLEEKQIALGLKEVKNEALNOVSOA 2297
QY 2122 SSVGPBK-AQOPOTAPSRAPPOPOPOSPAPEVQ--TQRPVQI-----2162
Db 2298 HSNNDVKIAENNGIAKISEVHEPILIKRNAKOEITQDAOSIDITMANNKSTNEKSAI 2357
QY 2163 -----QTVSSHVSEAPQTHAOS-----2182
Db 2358 DRVNAVAKIDAINITNATTTLQVMDAKMSGTISQILIPSTAVYTNALALASEAKKNNA 2417
QY 2183 ---KPYVAASQPOSN-----VQSGPYRVOSPS 2208
Db 2418 IIDQPNATAEKEEANKKVDRLQEADANILKAHTTEDEVNINKQAVONINAVQEVYIK 2477
QY 2209 QTRIRPSTPSQSPGQSOVOTTSQPIPIQPHISLOIPSGQOQSOQVQOSSQVOTLSSG 2268
Db 2478 KONVK-NQNLQFIDQKILINT-----PDATLEKKAANRLLOVNLSTDEILNV 2528
QY 2269 QTLNOVSYS-SPSRPOLQIQPOPOVIAVPOLQOOVLYLSIQSOVAVQ--IOAQOSGV 2324
Db 2529 DHNNEVDALDKARKIE-----ELVPOYSKKRDVLNMQEFNFNQTOLEIQENQAT 2580
QY 2325 PQQIKLQLPQIQOQ--SSAVQTHQIONVTVQAAVSQELQORVQOQLRDOQKKKQOQIE 2381

```

```

Db 2581 NEE-KTEALINKINQOLLNOQKAVNIDQOSNKNDVSAKTR-SIODIEQIQHPQK-----A 2633
QY 2382 IKREHILQASNOSEIIOQVAMKH-NAVTEHLKQKSMTPAEFEENQRIJCVQVKKYL 2440
Db 2634 TGRHRLNEKANO--QOSTIAHPNSTIE-----EKQESAKL--OEVLKKA 2676
QY 2441 DKIDKEKQAKRRKRESEVEQKRSKONATKLSALLFKHKEQLRAELKRLALDKDLOI 2500
Db 2677 AKIDKQOTNDVAKRYVNGIALENTILPATY-----KDKAKADVNAKE--QKNLQI 2727
QY 2501 EVOEELKRDIKIKKEDLMQLOAQTAVAPCPVTVLPAPRAPPPSPPPPVQVHTGL 2560
Db 2728 NSNDEATEEKLVAUSDNLNHNVEITNOAIEDADPTQOVNVE-----KXKGIG 2774
QY 2561 STPLPLPAVSOKRRKEEKSSSKKKMISTSKKKKDTK 2602
Db 2775 TIRDIQPLVVKKPPAKSKIESAVEKKKTETIQONATHDEV 2816

```

```

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PR
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

Query Match 2.58; Score 363.5; DB 4; Length 10182;
 Best Local Similarity 17.3%; Pred. No. 4e-12;
 Matches 573; Conservative 488; Mismatches 1236; Indels 1019; Gaps 142;

```

QY 136 VSOQCTLMAMHMYVLKAVLRBEDTSNTFFGPADLKDVSNTLYFTDGMTPEVLK-VY 194
Db 3117 IENOYNTAKNEAHNV-----ENTNPTVNAVVDALRKINA-----IOPEVTRKAIN 3161
QY 195 CESDKREYHNLVPOEADDPYGPVENKIKVLOFLVDOFLTNTIAREELMSEGVYQYDDHC 254
Db 3162 ILQKKEHNSL-----VRAREK-----LDQAINQSPSLNMGTOESINNTYTRK 3204
QY 255 RVCHKLDLCCETCSAVYHIECVKPPLEVEYEDMOCYCV-----AHKVPGVYDCV 307
Db 3205 REAQNINAS-----SADTIINNGDASIEQITTEKIVEEATNALNKAQHILADITSLK 3257
QY 308 AEIOKKNPYIRHEPIDGRSRRKTYFLNRLIIEEDTENENEKKIWTYSTKVOLAEILDC 367
Db 3258 TEVVK-----LSRR-----GDTNNKKPPSSVSNV-----NTIHS 3286
QY 368 LDKDYWEAE-----LCKILEMRKEEIRHMDITEDLTN-----400
Db 3287 LQSEITQTEHNANTLIINKPIRSVEVNNALCHEVNLQNLRLDTINLLQPLANRESLKEAR 3346
QY 401 -----KAGSKNSFLA-----NANEI---LESIR 422
Db 3347 NRLESKINETVQTDGMTQOOSVENVYKQAKIRQAQNSSIAQILINNGDSDDEVSTETELKN 3406
QY 423 AKGDIDINV-----KSPETEKKDKNETENDSDKAENREEFQDQLEK---DSDDKTP 472
Db 3407 QKLSELNINSHLTVNKEPLETAKNQLQANIDQKPSIDGMQOSVQSYERKLQEAQKIN 3466

```


0Y	473	D-----	DDPQKSEVGDGRKSEKSGELSESGAGK-----	ASGSRILTRLRNP	518
Db	3467	SINNVLANPDPVNAIRTNKVEBOINNELTQ----	AKQGLTVDPKLINAKTALOOSLDNQ	35233	
0Y	519	DSK--LSOLSKSOQVAAAHABANKLFEGKEVLVNSQGEISRLSTKRKEVIM--	CININY	574	
Db	3524	PSTGMTEATIQNTNAKROKAEVDIONAKI-IEANQPSVOQVSDPKSVEDALSELN-	3561		
0Y	575	FKIGOEKRYRVHNQYSTNSFALNKQHREDHDKRRHLAKFCLTPAGEFKWGSVHSG	634		
Db	3582	-----	AKSALRADKOELQO-----	AYNQLQPT-----	DLNKK 3610
0Y	635	VLTISTLRITLTOLENNIPBSFL-----	HPMA--	SHANNTKAVOMCSKREFAL	683
Db	3611	PASITANQORYOQSNELNSTKNTKTDRILKEQNPVADVNMLNMYREVQ----	QKLINE	36655	
0Y	684	ALALIECAVPRVYVLPJWREFLGHTRLHWTSTIEREKRYKKKEKKEO-----	EEDTM	737	
Db	3666	ARALLO-----	-----	NKEDNSALYAKREOLOAOVDVPSTGCM	36999
0Y	738	QOATWVRYTPPVKHQVAKQ-----	KGEERYVTVYGGGWSMIS--	KTHYVRVPLKPGNTVNY	792
Db	3700	TQOQT--KDDVYSKQQAQOEISKAQOVDINGDATTQOISNATNVERALHEL--	NAKTGL	37566	
0Y	793	R---KSLEGTKNNDEMDESDK-----	RKCSPPKIKITIEDSEKDEYK-----	835	
Db	3757	RADKEELONAYNLTQNTIDPISGKTPASIRKYNAKRSRIOTQIDASKNEANSILITNDNPQ	3816		
0Y	836	-----	SDAKGADQNDMDISKTTEKKD-----	DVKELSDSDSKP	871
Db	3817	SOVTPALNLIKAVOPELDKAIAMKKNENNALVQAKQOLOITVNEVDPTQGMTDTDTANN	3876		
0Y	872	CKEEPMVEYDDM--KTESHVACOSSQVDVYVNSEGHRTSYKTKKSSKGLDLEERRIK	930		
Db	3877	YKSKKREADEIQKAAQOIIINNGDATEGQITNEFRVYNAQIMINAKND-----	39252		
0Y	931	QPTLEEKORLEKILLEGITNG-IGKISTNSK--	NISEPVTITAKAKEGCSDSMKEQSP	987	
Db	3926	-----	LRAKSOLEAVYANQOLNIONVDYNGKPKASIOYQOAROAIE--	TOYNNAKSEHQ	39777
0Y	988	NANNDOP--EDLIGGCSQSSSVLRMSDPSH--	PTNK-----	LYPKDRYLDVSTIRSEPT	10388
Db	3978	ILLESNSPVNEVAQALOKVAAYOLKAVDAIHILQNKENNALSALVTAKNLOOSVN--	DQPLT	40366	
0Y	1039	KCPKONSTEN-DIEEKVYSDLASGQEPJ-----	KSXTKGNDFEIDSK--	1080	
Db	4037	TGMQODSINNEYEAKRNBQAGAIINAEVJINNGDATAKQISDEKSYEQALAILNNAKQL	40966		
0Y	1081	LASADIDIGTLI-----	CKNKKP-----	LIOESDITVSSSKS--	11333
Db	4097	TADTTELOTAVOQLNRGRDITNNKKPSINAYKAIOSLETQITLSAKDANNAVIGQPIRTV	41566		
0Y	1114	-----	LHSSVPKSTN-----	DRDAIPLSRADFEKGLG-----	11411
Db	4157	QEVNNALOQVNLQOJTEALINLOPLSLNNDALKAR--	LNLENKKNIQVTQDGMTQOSIE	4215	
0Y	1142	-----	CDESNSITLE--	NSSDTVSIOSSSEEDIVONSNESTISEQRTREDOVEYLE	11919
Db	4216	AYONAKKRAQNESTALALINNGDADBOQITTTEDVYNOQTT--	NLTQALINGITVAKPELE	4274	
0Y	1192	PLKCELVSGESTGNCEDRLP-----	VKGTENGKPKSQOQKLEERPVNKCSDQILKMTT	1246	
Db	4275	TAKTAL-----	QNNDIQVSTGDMTQOOSVANYNOKLOJAKNEIMTINN--	LANNP	43233
0Y	1247	D-----	KNNNEBESKKGQRTSTFQJNGKDNKPJITYLKGECLKEISESVSGNVEPK	1300	
Db	4324	DVNAIKTNKKEAERISNDLTQOAKNNLOV--	DTQPLEKIKRLODEIDOGTWTDMGTDS	4380	
0Y	1301	VNNIN-----	KIIPENDIKSL-----	TYKESARIRPINDVIMEDF--	NBRNS----- 13411
Db	4381	VDNTNDSIAAITLEKGVNKLNRNPTVDEQESV--	ANAOQVIOJQDNARISVLDPK	4436	

QY	1342	----	SBPSTKSHLLS-----	SDABG-----	NYRSLTLPSTHESOST-----	OTTPPS--A	1381
Db	4437	TQLOAKKARLENS	INQOTDIDGMTOQDSLNNYNDKLAKA	NONLEKTSKYIGOPVIAEIRQ		44966	
QY	1382	SCPESSNVAQVEDMEIETSEVKKYTSPTL-----	SEESNLSNDPIDEJENLPINK	14322			
Db	4497	NTDEMAIKKQALD-----	TARQQLTLINEPIFINIINSHESLNMA	NONKONFEAQNVS	45466		
QY	1433	NEENVGESKRTVITEVYTMSTVAEESKTVIKVE-----	KGDQOTVASTENC	1481			
Db	4547	APNHN-----	TLETTIKNKADTILNOSMTALSESIA	DYENQOQOEYVLDA	SNKKRQDYDNAVNA	46030	
QY	1482	AKSYTTTTTTTTV-----	TKSTPSTGSDIISVGEQSKTVYTT-----	1520			
Db	4604	AKGILINQOEPSPTMSADVIDOKAEDVYKTKTALDGNORLEVA	QAQALNHLNTLINDLNDOR	4663			
QY	1521	-TVTDSLTTT-----	-GGTLVTSMTVSK-----	YSTPDKVKLAKF	1554		
Db	4664	QTLIDTINHSPININSVQAQEKANTVATANTQLOKOTI	ANYDDELHDGNTINMDK-----	4717			
QY	1555	SRPKTNSGTALPSYRKFVYTKSTKSLIYVLPNDLKKLAK-----	GGIREVYFNY	1606			
Db	4718	-DKDAVNANVA	NAKOLINDSANOQALDP--AEINKYQVRYVTKNDLNGNDKLAEAKR	4774			
QY	1607	NAKRALDWPYRPSRPFPGTITWKRYRQUTVYKSLAGVSLMLRLMASLRMDMAKAP	PPGCG	1666			
Db	4775	DANTTID-----	-GLTYLNEAQRN-----	AKENVKA	4801		
QY	1667	STPTE-TSETE-----	ITTYELIK-----	RBDVGP--YGIREFYCIARK	1701		
Db	4802	STKRTITSOLODYNO	LIMMALRNSVNDYNNVANKASNTINEDNGPKREYNAOAVIHA--QT	4866			
QY	1702	IICPIGVETREKETPTPQORKGLSS-----	ALRPKRETPPKQTG--PVIIETWAE	1751			
Db	4861	LINAQSNPEMRDQVNOKTOAVNTAHONHGLHQOKLEQAQSSANT	EIGNPLNTINQOAKE	4920			
QY	1752	ELELMEIRAFERERKEKAQAVEQAKRLEQOKPVIATSTTSPSSSTTISISQAQVM	18111				
Db	4921	K-ELVNSQOTPEVOEOLNQA-----	KSLDSSMGTLISLVAKOPVOKTQSYVINE	DPDE	4973		
QY	1812	VAPISGSVYTTGKVLVLTTKVGP-----	AVTFEQONKNFHQFATVWKOQOSNS	1860			
Db	4974	QSAVNDSTITTMQOTI--	NKTRADVPLODKTLVDNIAINSITREKNA	LHGEOKLTTAKTEALINA	5033		
QY	1861	-----	GVVOVOOK-----	VLGIIPBSSTGTSOQFTSFOPRATVITRPN	TSGSGGTTSSNQ	19111	
Db	5032	INTLADLNTPOKEAIKTAINTAHTRTDYTAEGSKANOINSAMH	TLTRONISDNESVTNESN	5091			
QY	1912	VITG-PQIRPQMT-----	VIRTPLOQSTL-----	GKAIT-----	1939		
Db	5092	YINAEPEKOHAFETALNNAKEIYNE--	QOATILDANSINOKAOLITTKNALDGEOLBRR	5149			
QY	1940	-----	PRPVWVGAPPOVY-----	MTQILIRG--	1966		
Db	5150	KENADQELINTLNQLTQARNSSEGLINSSOTRIEVA	SOLAKAEKELMKWEOJNLHILNCKN	5209			
QY	1961	-----	OPVSTAASAPRTVY-----	STPFGOKSLTSA	TSTNSIQSSAQP	2001	
Db	5210	QMINSKRTINEDANOQOAYINALIASALKNKQONELDKVITIEQA	LINNINSAINN-----	5265			
QY	2002	QOGQVKLTMAQ-----	LTOJLTQGHGNGOGLTVIVIOGOGQOTTGOLQILPQGVTVLP	2052			
Db	5266	INGEKILTKAKEDAVASINNIISGLTINBOKTERNOAV--	NGAOTRQOAVANKLRDAEALDQ	5322			
QY	2053	PGOOLMOAAMNGVQAF--	LFTPLATATTA	STTTVYSTTAAGIGEDROSKLSPOMQV	2110		
Db	5323	SMQTLRDLVNNQNAIHSTSNYFNE	DSOTKNTYDMAIDNSSTYITGOHNE	LKSTIDQTI	5382		
QY	2111	HODTLRPAOSSVGP	AKAOPOTAPASBPQOTQOSPAQOEVOPTOPVQOTVY--SS	2168			
Db	5383	SRINT--	AKKMDLHGVCRLORDKG-----	TANQELGOLGTYLNDPOKSEEBLVNGS	5430		
QY	2169	HVPSEADPTHAOSSKPOVAASQPOSNWQOSDPVRQSPOTRIRP	STPSPOLSPGQOQSV	2228			


```

Db 5431 NTRSEVE-EHLNKAESLNNNAKOLRDKVAEKTNYK---QSSDYINDSTEHORGVDOLQOE 5486
QY 2229 OTTSQPIPIPHTSLOIPSGGPOQ-SOPVOSSQTQLTSSQT----- 2270
Db 5487 AENINEL-----GNPTLNSEIEOKLQOLTDQNALQSHLLEAKNNAIT 5533
QY 2271 -LINVSSSPSPOLQIOPPOVIAVPOLOQOQVLSQIOS--QVVAIOIQAQSGVPQ 2327
Db 5534 GINKLTALNDARQKALENVQAQO-TIPAVNQQLLDREIFTAQAOLRDKQAQNNVHQ 5552
QY 2338 I-----KIQLPFIQIOSSAVOTHOIIONVTVQAASVDEQL 2362
Db 5593 SNYNEDEOPKHNVDNSVQAQOTIIDKIQDPIMKNMETEQAINQINTQT--ALSGENKL 5650
QY 2363 QRVOLRDOQ---QKKQOQIEIKREHTLQASNOSEIQAQVMMKH-NAYIEHLKO--- 2414
Db 5651 HTDESTNRQLEGSLNTAOTINAEKDLVNOAKTRIDVQAQLAAKEINSMSLSRQIO 5710
QY 2415 -----KKS-----MPAREENORMIVCNOVMKYILDKIDKEKQAARRRE- 2457
Db 5711 NKEDIKRSSAYINADPTKVTAYDQALQNAENIINATPVELNKATIEDQALSROVQAQODL 5770
QY 2458 ESVEQ-KRSQONAK---LSALLFKHEQRLAEILKRALLDLOIEVEELKRLDKLT 2512
Db 5771 DGVOQLANAKQAOTQVNGLSLNDGOKRELNT-----LINSANTRTKVOEELNKATEL 5824
QY 2513 KKEEDLM-----QLAATAVAAPCP-----VTPVL----- 2539
Db 5825 NHAELARNSQVNDYQKQSSNYNVEDQPEGHYNDANVNEQAATINNAPVLDKLALER 5884
QY 2540 -----APPAPPPPPPPPGVOHTGLSTPTLPVSAOKRREKSSSKS 2585
Db 5885 LVTQVNTTKDALHGAOKLTQDOQAETGIR--GLTS-----LNEPKNAEVAKVTAT 5936
QY 2586 KKKIISTSKETKDDTLGYCCTPYDE-----SKFYIGDROQNMWGHGRCVGLQSEAE 2640
Db 5937 RDEVRNRQETTLIDTLMGLIKRSIKKDKNTKNSKTIINEDHQQAQVDAVNAQO---Q 5993
QY 2641 LIDEEVPCQOSTEDAMTVLPLT-----EKDEYGLKRYLSLOAKHMAPE 2687
Db 5994 VIDE---TQATLSGDTINOALANAVTQAKSNLHGTKLQHDKDSKQTLAQOLNNSAQKH 6050
QY 2688 LEVPDPADADYGVIEKPDLMERVOARYE-----KLEFVADMTKIPDNC 2738
Db 6051 ME-----DSLIDNESTRTOVQHDLEAQAQDLGALKAKESIKDYTNIVSNG 6096
QY 2739 RYVN--PSDSPFYQCA 2752
Db 6097 NYINAEPSKKQAYDAA 6112

```

```

RESULT 5
US-08-046-585-5
; Sequence 5, Application us/08046585
; Patent No. 5453362
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Hett, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-Apr-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-046-585-5

```

Query Match 2.5%; Score 353; DB 1; Length 2035;

Best Local Similarity 21.4%; Pred. No. 1.7e-12;

Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

```

QY 1368 TKESDSQOTT-PSASCPESNVQVEDMELEISEVKKV-----SSPTSEESLSN 1420
Db 619 TSVSSAINTSTRPIITVHKSGVTVAAQAOVTVTVGAVTKTTLVKSPTSVSGSAL- 676
QY 1421 DFIDENGLPIKNENNGESKRKTVITEVMTSTVATESEKT-----VIRVE---KGDR 1471
Db 677 -----ISNLGKVMSSVQTKRPVQSAVIGQASGTGVVQIIOITKGPLAGTI 721
QY 1472 QTVVSSTEENAKSTVTTTTVTKLSTPSTGGSVDIISVEQSKTVVTTVTS----- 1525
Db 722 LKIVTSAD--GKPTIITITTOASGAGTKPT--ILGISVSPTTKPQTITTIITIPMSAI 777
QY 1526 LTTTGGTLVMSMTVSKESYTRDKVKLKKFSRPKTRSGTALPSY-----RKFTKSTKKS 1580
Db 778 ITQAGATGVSSSPGKIKPTIITTKVTSSTGAPAKITTVAPRIANGHQGQGVTVYVKG 837
QY 1581 IFVLPNDLKLKLRKGGIREVPYFNNA--KRALDIWYPSRPTFGITWRYRLQTVKSLA 1639
Db 838 APQPGFTLRTVP--MGVRLVTPVTSVAKPAVTVLV---KGTGVT--TLGTVTGIV 890
QY 1640 GVSIMLRLLMASLRWMDMAKVPFGGSGTSTSEHTTEITTEILKRRDVGPYIREYCI 1699
Db 891 STSL-----AGAGHSTVASLAPITTLGTITALS----- 920
QY 1700 KLIICPIGVETPKETPTPOKGLRSSALRPKRPTPKQ---TGPVILETVAEELET 1755
Db 921 SYVINPALTIVSAQOTTLTAAGLITPTITMGOVSPQTOVTLTASGV- 970
QY 1756 WEIRAFERVEKEKAQAVEQAQAKRLDQOKPVYIANSTISPTSSSTISPAQVAVAPI 1815
Db 971 -----AQPV-----HDLPSIILASPTTEQPTATVTTIADSGGDVQD- 1006
QY 1816 SGSVT-----TGKMWLTTKV-----GSAATVTFQONKMFHQTFAPVWQ--G 1856
Db 1007 -GTIVTLVCSNPPCETHETHTGTTNATTVVANLGGHPQPOVQVCDROEAAASLVSTVG 1065
QY 1857 QSNQGVVOQKVLGIIPSTGTSOQFTSFQPRATVTVIRPSTSGGTTSSQVITGP 1916
Db 1066 QQNGSVVRCNSN-----PCET--HETGTTNATTAT-----SMAAQHGCNS- 1106
QY 1917 QIRPGMTVIRTPLOOSTLGAIRTPVWVOPGAPQO--VMTQIIRGQPVSTAVS-APNTV 1973
Db 1107 -----PCETHETHTGTTNATTAMSSVGANHORDARRACAAGTPAVIRISVATGAL 1156
QY 1974 SSTPGQKS-----LTSATSTSN-----QSSASQPP-----RPQOG 2004

```


Db 1157 EAAQSKSQCOTROTATSTMTVMAATGAPCSAGPLIGPSMAPEPGRSFAVQLAPLSS 1216
 QY 2005 QVLTMAQLTQLQGHGNGGLVVIQGGQOTGOLILP-----OG-----VTVL 2051
 Db 1217 KVALSSPSIKDLPAGRHSHAVSTAAMTRSSVSGEPRMAVCSLOGGSTVVTYTALE 1276
 QY 2052 G---PGQQLMQ-----AAMPN-GTVQREFLPLATTAATTAATTTTTS 2090
 Db 1277 ALLCPATVTVQVCSNPCEHETHTGTNTATTSNAGSAQVRCNSNPCETHETGTHATTA 1336
 QY 2091 TTAAGGEBOROSKLSL---OMOVHODKTLPPAOSSSVG-----PAKAPQV-----AQ 2135
 Db 1337 TSNAGTGQPEGGQPPAGRPCETHHTGTSTGTMSVSVGALLPDAATSSHRTVESGLEVAAA 1396
 QY 2136 PSARPQPPQPOPS-----AOPEVOTOP--EVOTOTVSSHVPSAPQTHAOSKPOV 2186
 Db 1397 PSTTPQAGTALLAPFTQVNCNSNPCETHETGTHATTAATTSNAGSNDPPRPAASDQGEV 1456
 QY 2187 AAOQSPQSVQGSFVAVOSPSQ---TRIRPSTP---SGLSPGOOSQVOTTSQPIPI 2238
 Db 1457 ESTQGSVNITSSAITTTTSTLTFRATVTVTQSTFVPGSPVPPPELQVSPGRQOLP- 1515
 QY 2239 QPHTSLQISQGGPOSGPOQVOSTOTLSSQOTLNOVSVSPSRPQ-----LQIQ 2288
 Db 1516 -PQQLQASALMGEAEVLSASQTPPELPAVADLSSTGEPSSGQBSAGSAVAVATVWQP 1574
 QY 2289 POPVIAVAPQLOQOVVLSQIOSOV-----AOIQAGSGVPOQIK 2329
 Db 1575 PPPTQSEVQLSLPQELMAEAQAGTTTMTVGTLPPELAVTAAEAQAATAEBAQALA 1634
 QY 2330 LQLPQIQSSAVQTHQIQNVTVQAASVQEQLOQVQQLRDQOKKKQOOIEIKREHTLQ 2389
 Db 1635 IQAVLQAAQAAVAGTGEPMDSQAATVQAEIGHL-SAGQGGQATTPIVLTQOEIAA 1693
 QY 2390 ASMQSEILQKVVMKNAVIEHLKOKKSMTPAE 2422
 Db 1694 LVQOQQLQEAQAOQH-----HHLPTEALAPAD 1721
 RESULT 6
 US-08-393-703-5
 ; Sequence 5, Application US/08393703
 ; Patent No. 5585239
 ; GENERAL INFORMATION:
 ; APPLICANT: Lamarco, Kelly
 ; APPLICANT: Wilson, Angus
 ; APPLICANT: Heirt, Winship
 ; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/393,703
 ; FILING DATE: 24-FEB-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249

; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2035 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-393-703-5
 Query Match 2.5%; Score 353; DB 1; Length 2035;
 Best Local Similarity 21.4%; Pred. No. 1.7e-12;
 Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;
 QY 1368 TRKSDSQTTT-PSASQPESSNVQVDEMEIESEVKVY-----SSPTSEESLSLN 1420
 Db 619 TSVSSATNTSTRIITYHKSGIVTVAAQAOVTVTVGGVTKTTLVKSPISVGGSL- 676
 QY 1421 DFDENGLPIKNKNVNGESKRRKTVITEVTMTSTVATESKT-----VIRVE---KDK 1471
 Db 677 -----ISNLGKVMSSVQTKPVQTSAVTGAQSTGPVNIQTKGLPAGTI 721
 QY 1472 QTVVSTENCAKSTVTTTTTTLKLPSTGSGVDLIISVKEQSKVTVTTVTDS----- 1525
 Db 722 LKLVTSAD--GKPTTITTTQASGAGTKP--ILGISVSPSTTKPQGTITTIIRTPMSAI 777
 QY 1526 LTTTGGTLVTSMTVSKREYTRDKVTKLMKFSRPKKTRSGTALPSV-----RKFTKSTKS 1580
 Db 778 ITQAGATGVTSSPGIKSPIITITTKVWFGSGTGAPAKITIAVPAIANGGQGVTVQVYKLG 837
 QY 1581 IFVLPNDDLKLARKGIREVPEYFNNA-KPALDIWYPSRPFTGITWYRLOQVYKSLA 1639
 Db 838 APQPGTILKTV-MGVRLVTPVTVSAVKAVATTLV--KGTQVTV--TLGTVTGV 890
 QY 1640 GVSLMLRLMASLRWDDMAKVPKGGSTRETSETEITTEITLIIKRQDVPGIREFYCI 1699
 Db 891 STSL-----AGAGHSTYSASLAPPTLGTATLS----- 920
 QY 1700 RKIIICPVPEPKETPTPOKGLRSSALRPKRPERPKO---TGPIIETWAEEL 1755
 Db 921 SQVINPFAIVSAQITTLTLAGGLTPTTMOQVSOPTQVTLTADSGV----- 970
 QY 1756 WEIRAPAEVKEKAQAVEQAKRLQOKPVIATSTSPSTSTSPSPAOKVAVAP1 1815
 Db 971 -----AQV-----HDLPSILASPTTEQPLATVITADSGQGVQ- 1006
 QY 1816 SGSVT-----TGKMYLTKV-----GSPATVTFQONKNFHQTFATWVKO--G 1856
 Db 1007 -GVTTLVCSNPCEHETHTGTNTATTVVANLGHQPPQVQVCDPQOEAAASLVSTVG 1065
 QY 1857 QNSGVSQVQKXVLGIIPSTGSOQTFSEFQPRATVITIRPMTSGSGGTSNSQVITGP 1916
 Db 1066 QQNGSVYKVCNSP---PCEV---HETGTTNTATTAI---SNMAGOHGCSN----- 1106
 QY 1917 QIRPGMTVIRTPLOQSTLGAIRTPVMVQPGAPQ--VMTQIIRGQPVSTAAS-APNV 1973
 Db 1107 -----PCETHETGTTNTATTAAMSSVGANHQDARACAGIPNIRISVATGAL 1156
 QY 1974 SSTPGOKS-----LTSATSTSN-----OSSASQPP-----RPQG 2004
 Db 1157 EAAQSKSQCOTROTATSTMTVMAATGAPCSAGPLIGPSMAPEPGRSFAVQLAPLSS 1216
 QY 2005 QVLTMAQLTQLQGHGNGGLVVIQGGQOTGOLILP-----OG-----VTVL 2051
 Db 1217 KVALSSPSIKDLPAGRHSHAVSTAAMTRSSVSGEPRMAVCSLOGGSTVVTYTALE 1276
 QY 2052 G---PGQQLMQ-----AAMPN-GTVQREFLPLATTAATTAATTTTTS 2090
 Db 1277 ALLCPATVTVQVCSNPCEHETHTGTNTATTSNAGSAQVRCNSNPCETHETGTHATTA 1336
 QY 2091 TTAAGGEBOROSKLSL---OMOVHODKTLPPAOSSSVG-----PAKAPQV-----AQ 2135
 Db 1337 TSNAGTGQPEGGQPPAGRPCETHHTGTSTGTMSVSVGALLPDAATSSHRTVESGLEVAAA 1396

QY 2136 PSARPQOTOPSP-----ADPEVOTOP--EVQOTVTVSSHVPSPAOTTHAOSKMPV 2186
 Db 1397 PSVTPQAGTALLAPFPYQVRCVSNPCETHETGTHATVTSNMSSQODPPAPASDQGEV 1456
 QY 2187 AASQSPQSNVQOGSPVAVQSPSQ-----TRIRPSTP---SOLSPGOOSQVQTTTSPQIP1 2238
 Db 1457 ESTQGDVNTITSSAITTTVSTLTTRAVTVTVOSTVPVPGSVPPPELQVSPGROQLP- 1515
 QY 2239 QPHTSLQIPSGQOPQOSPOVOSSSTOTLSGOTLNOVSVSSPRQ-----LQIQ 2288
 Db 1516 -PRQLQASASTALMGESAEVLASQTELPAAVDLSSTGSPSSQESAGSAAVAVTVVQ 1574
 QY 2289 POPQVIAVPOLQOOVQVLSQISQV-----AQIQAOQSGVPOQIK 2329
 Db 1575 PPPQSEVVDLSLPQELMAEQAGTTTLVNTGLTPRELAVTAAEAQAATEEAQALA 1634
 QY 2330 LQLPQIQIOSSAVOTHQIQNVTVVQAASVOEQLQVQOQLRDOOKKKQOIEIKREHTLQ 2389
 Db 1635 IQAVLQAQAQAVMGTEGPMPTSEAAATVTAELGHL-SAGQEGQATTPIVLTQELAA 1693
 QY 2390 ASNOSEIIQKQVYMKHNAVIEHLKKOKSMTPAE 2422
 Db 1694 LVQOQQLQEAQAQOQH-----HHLPTALAPAD 1721

RESULT 7

PCT-US93-11721-5
 ; Sequence 5, Application PC/TUS9311721
 ; GENERAL INFORMATION:
 ; APPLICANT: Lamarco, Kelly
 ; APPLICANT: Wilson, Angus
 ; APPLICANT: Herr, Winship
 ; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
 ; TITLE OF INVENTION: HOST CELL FACTOR
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; City: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/11721
 ; FILING DATE: 03-DEC-1993
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: PP-57503-1/RAO
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2035 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MODE/OLE TYPE: peptide
 ; PCT-US93-11721-5

Query Match 2.5%; Score 353; DB 5; Length 2035;
 Best Local Similarity 21.4%; Pred. No. 1.7e-12;
 Matches 264; Conservative 146; Mismatches 515; Indels 308; Gaps 48;

QY 1368 TKESDSTQTTT-PSASCPESNSVQVEDMEIETSEVKKYV-----SPTTSEESNLSN 1420
 Db 619 TSVSSATNTSTRITITVHNSGTVTAQAQAVYTVVGVGKTKITLTKSLISVSGSAL-- 676
 QY 1421 DFIDENGLPINKNEVNGESKRRTVTEVTMTSTVATESKT-----VAKVE---KGK 1471
 Db 677 -----LSNLKVMASVQTKRVQTSAAVTGAASGVPVQIIOQTKPPLPAGTI 721
 QY 1472 QTVVSTENCAKSTVTTTITVTKLSTPSGSGVDIISVEKQSTVAVTTVTDSS----- 1525
 Db 722 LKLVTSAD--GKPTTITITVQAAGAGTKPT--ILGSSVSPSTTKRGCTITITITMSAI 777
 QY 1526 LTTTGGTLVTSMTVSKREYSTRDKVKLTKSPRKTSGALPSY-----RKEVTKSTKS 1580
 Db 778 ITQAGATGVTSISPGIKSPIITITTKVTSQSGAPAKIITAVPKIATGHGQGVQTVVLGK 837
 QY 1581 IFVLPNDDLKLLARKGIRVVPFNNNA-KPALDIMPSPRPTFGITRKYRLQTYKSLA 1639
 Db 838 APQOPGTILKTVP--MGVRLVTPVTVSAVKAQVATTVLV--KGTGTGVT--TLGVTGTV 890
 QY 1640 GVSILMLRLMASLRWMDMAKVPFGGSGSTRTESEITETTELTKRRDVGPGIRREYCI 1699
 Db 891 STSL-----AGAGHSTASALAPITTLGTLATLS----- 920
 QY 1700 RKIICPIGVETPEKETPTPQKGLRSSALPKRREPQK---TGPVLIETVWAELEL 1755
 Db 921 SQVINTPAITVSAQAOTTLTAAGGLTPTITMQVSPQTVTLITARSVE----- 970
 QY 1756 WEIRAFAREYKEKAQAVEQAKKRLQOKPYIAISTSPSTSTISPAQKVVAVPI 1815
 Db 971 -----ADPV-----HDLPSILSPTEOPATVTLADSGQVQDP- 1006
 QY 1816 SGSVT-----TGKRVLTTRV-----GSPATVTFQONKNEHQTFATVQK--G 1856
 Db 1007 -GVITLVCSNPPCETHETGTTNTATTIVVANLGHQPPQVQVCCROEAAASLVISTYG 1065
 QY 1857 QNSNGVVQOQKVLGIIPSTGSOQTFSTFORPATVIRTPSSGGTTSQVITGP 1916
 Db 1066 IQMGSVYRVCSNP---PCET---HETGTTNTATTAT---SNMAQOHGCSN----- 1106
 QY 1917 QIRPGMTVIRTPLOQSTLGAIIIRTPVWQPGAPQO--VMTQIIRGQPVSTAVS-APNVY 1973
 Db 1107 -----PPCETHETGTTNTATTAMSSVGNHQAQDARACAAGTPAVIRISVATGAL 1156
 QY 1974 SSTPGOKS-----LTSATSTSNL-----QSSASQPP-----PPQG 2004
 Db 1157 EAAQGSKSQOQTRQTSATSTFTMTMATGAPCSAGPLGPMAREPGGRSPAFLPLSS 1216
 QY 2005 QVKLTMAQLTQLQGHGNGGLVYVIOGOCOTIGQLLP-----VTVLP 2051
 Db 1217 KVRLLSPSINDLPAGRSHAVSTAAMTRSSVGAQEPMAVCESLQGSSTTVVYTALE 1276
 QY 2052 G---PGQQLMO-----AAMPN-GTVQRFLETPLATTTATTTATTTTTS 2090
 Db 1277 ALLCPSATVTVQVCSNPPCETHETGTTTATTSNAGSQAQRVCSNPPCETHETGTHATT 1336
 QY 2091 TTAAGTEGROKSLSP---OMQVHQDKTLPPAOSSVG-----PAKAQPT-----AQ 2135
 Db 1337 TSNQGTQPRGGQQPPAGRCPCETHQTTSTGTMSVSGALLPADTSSHRVESGLEVA 1396
 QY 2136 PSARPQOTOPSP-----ADPEVOTOP--EVQOTVTVSSHVPSPAOTTHAOSKMPV 2186
 Db 1397 PSVTPQAGTALLAPFPYQVRCVSNPCETHETGTHATVTSNMSSQODPPAPASDQGEV 1456
 QY 2187 AASQSPQSNVQOGSPVAVQSPSQ-----TRIRPSTP---SOLSPGOOSQVQTTTSPQIP1 2238
 Db 1457 ESTQGDVNTITSSAITTTVSTLTTRAVTVTVOSTVPVPGSVPPPELQVSPGROQLP- 1515
 QY 2239 QPHTSLQIPSGQOPQOSPOVOSSSTOTLSGOTLNOVSVSSPRQ-----LQIQ 2288
 Db 1516 -PRQLQASASTALMGESAEVLASQTELPAAVDLSSTGSPSSQESAGSAAVAVTVVQ 1574
 QY 2289 POPQVIAVPOLQOOVQVLSQISQV-----AQIQAOQSGVPOQIK 2329

Db 1575 PRPTOSEVNDLSLPOELMAEAQAGTTTLAVTGLTPEELAVTAAEAQAQAAATEEACALAA 1634
Qy 2330 LQLPPIQOOSAVOTHOIQNVTVQAASVOEOLQRYOOLRDQOQKQOQIEIKREHTLQ 2389
Db 1635 IQAVLQAQAQAVGTEBPMDTSEAAATVYQALGHL-SAGQEGQAATTPIVLTQOELAA 1693
Qy 2390 ASQOSETIQKQVVMKHNVAIEHLKOKKSMTPAE 2422
Db 1694 LVQOQOOLQEAQAQOQH-----HHLPTLEALAPAD 1721

RESULT 8
US-08-821-355A-7
Sequence 7, Application US/08821355A
Patent No. 5851775
GENERAL INFORMATION:
APPLICANT: Barker, Nick
APPLICANT: Clevers, Hans
APPLICANT: Korinek, Vladimir
APPLICANT: Morin, Patrice
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
APPLICANT: Sparks, Andrew
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
TITLE OF INVENTION: Interact to Prevent Cancer
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,355A
FILING DATE: 20-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UT
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5851775e
US-08-821-355A-7

Query Match 2.4%; Score 338.5; DB 2: Length 2973;
Best Local Similarity 17.1%; Pred. No. 2.2e-11;
Matches 431; Conservative 337; Mismatches 904; Indels 845; Gaps 103;

Qy 387 EIHHRMDITEDLTNRKAGSNKSFLLA-ANEELLESIRAKGDIDNVKSPDEETKDN--- 442
Db 802 DYNRRDDNRSDFNMT---GNMTVLSPYLMTVLPSSSSSSRGLSDSSRS---EKDSLSLR 854
Qy 443 -----ETENDSKAEKKNREFEPOS-----LEKDSDDKTDYDDPEOGK----- 480

Db 855 ERGIGLGNHYPATENPETSCKRGLQISTTAQIAKWEVYSAIHTEOEDRSSGSTEHLHC 914
Qy 481 -----SEVGDf-KSEKSN-----GELSESPGAG 502
Db 915 VTDERNALRRSSAAHHTSNNTYFTKSSNSNRICSMYPAKLEYRRSSNDLSNVSSTGCG 974
Qy 503 KGASGSTRITITRLRNPDSKLSQLKS--QVAAAHAANKI-FKEGKEVLVN----- 551
Db 975 KRCQMKPISIESYSEDESKFCSYQYPADLAHKIHSANHHDDMDGELDPTINYSLKYSDE 1034
Qy 552 --SOGELSRISTKKEVYIMKGNINNYFRLGEGKRYVHNQYSTNSFPLNKHQREDK- 608
Db 1035 QLSNGRQSPQONERMAPKHIIIEIKOSEOROR--NOSTT--YPVYESTDDHLKF 1089
Qy 609 RRLAHKFLCTPAGFEFKWNGSVHGSKVLITSLRLTITOLENNIPSSFLHPNNAHRAW 668
Db 1090 QPHFGQOECSYPRSKRANS-----ETRRVGS-----NHQINQ 1123
Qy 669 IKAQMCSPREFALALALIECAVKPVMLPIWREFLGHTRLHMTSIEEREKVKKE 728
Db 1124 NVQSILQED-----QHEEERPTNY-----DYEDDKPTNYSE 1145
Qy 729 KQDEEETMQATWVKTTPVKKQWKKQGEYRVYGYGSMNISKTHYRFPKPLRGNT 768
Db 1146 RYSEEE-----QHEEERPTNY----- 1162
Qy 789 NVVYRKSLEGTKNMNMENDES-----DKRCSRPKIKIEPDEKD 831
Db 1163 STYNEE---KHHVQPIDYSLKYATDIPSSQKSFSSKSSGGSSTKEHSSSENT 1218
Qy 832 EVKSDAARGADONEMDISKITEKDDVYKELSDS-----DKP-CKEEPM 877
Db 1219 STPSSNAKR---QNLHPSSAQSRGQPKAATCKVYSINQETIQYCVEDPTICSRCS 1275
Qy 878 EYVDDDKTESHVNCQESSQ-----VDVYVNSEGHRLT-----STKKTSS 919
Db 1276 SLSSLSAEDIEICNOTQDASANTLOIAIEIKIGTRSAEDPVSEPAVSQHPRTKSS 1335
Qy 920 KLDGLLERRIKOFTLEKORLEKIGIKIGIKSTNSKN-----LSESPVI-TKAK 973
Db 1336 RLOG-----SSLSESAHKAVERFSSGAKSPSKGAQTPKPEHYVDETILMSRGT 1388
Qy 974 EGCQSDSMRQESPMANNDQPEDLIQCCSQSDSVLRMSDPSTHTKLYPKDRVLDVSI 1033
Db 1389 SVSLDSFESRSIASSVQBP-----CSGMVSGIISPDLPSDPQTMPPSR----- 1435
Qy 1034 RSPETKCPKONSIENIEEVVSDLASRGQPTKSKTGNDFIIDDKLSAD----- 1085
Db 1436 --SKTPPPPTQATKRE-----VPKNAKPTAKRESG-----PKOAAVNAVQRYOV 1481
Qy 1086 --DIGTLI-----CKNK-----KPLQIEESPTIV-----SSKSALHSSV 1118
Db 1482 LPADTLLHATSTPDGSCSSLSALSIDEPTIOKDELRIIMPVQENDKNGNESEDQ 1541
Qy 1119 PKSTNDRDATPLSRAMDFEGLCGDSESNSTLENSDVTYSIODSEEDMIVQNSNISIE 1178
Db 1542 PKESNENQ-----EKEAEKTIIDSEK--LDDSDDD----- 1570
Qy 1179 QFRTRODVEVLEPLKCELVSGESTGNCEDRLPYKTEANGKAPSOQKLEBPVKKCD 1238
Db 1571 -----DIEILE--EC-IISA-----MPTKSSR-KAKKPAOTASKLPPVARKPS 1610
Qy 1239 QI-----KLKNTTOKKNNENRESEKKGQRTSPFOINGDNPKIYLKGECLKEISESVVSG 1295
Db 1611 QLPVYVLLSQNLQFOQKHVSFTPG-----DDMPRYV--CV-----EGTPIHF 1651
Qy 1296 NWEPKVNNINKIIPENDIKSLFYKESAIRPFINGDIYIMEDFERNSSFTKSHLSSDAE 1355
Db 1652 STATSLSDLTIESPPNEL-----AAGEGVGAQOGEFEKRDITPIEGR--STDENQ 1701
Qy 1356 GNRDSLETLFPTIKES-----DSTQTTTPASCP 1384
Db 1702 GGTSSV-TIPELDNKAEEGDIACINSAMPKGSRRPFVYKIMDOVOQAASASSAP 1760

Qy	1385	ESNSVNOVEDMEIETSVKAVTS-----	SPITSEESNLSN	1420
Db	1761	KNKNDLCKR--KKPTSPVKPIPONTETPRVRKRNADSKNNLNAERVSDNKKDKÖNN	1818	
Qy	1421	DFIDENGLPIKNKENVNGE-----	SKRK	1443
Db	1819	NSKPFNKLNNEDNRVAGSFAFSPHPHYRIGETPRCPFSRNLSLSDFDODDVLSEK	1878	
Qy	1444	TVITEVTMTSYAVTESKIYIAKEKDKQTVVSTBENAKSTVY-----	TTTTV	1493
Db	1879	---AELKRAENKESKAVTSHELTLSNOOSAKKTOAIKOPINRGPRILÖKOSTPQÖ	1935	
Qy	1494	TKLSTPGSGVDIISVKEQSGKTVVTVTVTDSLTGTVLSMTVSKYSPRDKVLMK	1553	
Db	1936	SSKPIPRGAAFD---EKLÖNFALIENTPVCFS---HNSLSLSDIDÖENNKENPFIKE	1989	
Qy	1554	FSRP-----KTRSGTALPSTYRK-----EVTXSKKSIFVLPNDL-----	1569	
Db	1990	TEPDSOGESKPPQASVAPKSFHVEDTVPCFSRNSLSLSLSDIDDLÖECTISSAMPK	2049	
Qy	1590	-KTLARKGGRREVYFVYNAKP-----ALDIWYPSRPRTPGT-----WRY	1630	
Db	2050	KKKRSRLKÖDN-----KSPNNMGILGEDLTDLKIQÖRPDSEHGLSPSENFDMKA	2103	
Qy	1631	RLÖTVKSLAGVSLMLRLMLASLMDMAAKVPPGGSTRETSETEITTEITIKRDYGP	1690	
Db	2104	IOEFANSIV-----SLHQAAAACL-----SKQASDSÖD---STLSKS---	2140	
Qy	1691	YGINFEVCIRKIICPIGVPETPKETPTPQKGLRSSALPKRPPEPKGPIVIEITWAE	1750	
Db	2141	-GI-----SLGSPHLPDPÖEKPTSKNGPR--ILNPGKST-----LETKKIE	2182	
Qy	1751	EE-----LELMETRAFAERVEKEKAQAEQÖAKRRLÖQKP-----TVAT	1791	
Db	2183	SESKGINGKAKVYKSLITGKVRNSÖ-----ISGÖMÖPLÖANMPSISRGTMHI	2233	
Qy	1792	STPEPTSTSTISPPAKVNAVAPISGVYTTGTMVLTTTVGSPATVYTPÖÖKNFÖHOTAT	1851	
Db	2234	PGVNSSSSSTSPVSKKPPKLTSPASKSPSGÖATTISPPGAPSV-----	2278	
Qy	1852	VWKGÖSNGSVVOQÖVLGLIPSSGTSGÖQFTSPÖPTAVTVTRPNTSGSGGTTNSQÖ	1911	
Db	2279	-----KELSPVARKÖTSQIG-----GSSK-----APSSSGSRDSTPS-	2310	
Qy	1912	VITEPÖIRPGMTYIRPPLQÖSTLGAKAIKTPVWVOG-----APÖOVMTÖIRGÖPVSTAV	1967	
Db	2311	-----RPÄQÖPLSRPIÖSPGRNS-----ISPGRNGISPPKMLSPPTSSPSTA-	2354	
Qy	1968	SAPTVASPTGÖKSLS-----ATPSNIOSSASÖPPRPGÖQYKLTMAÖLTÖLTQÖ	2018	
Db	2355	---STKSSGSKMSYTPGÖMÖQÖNLTKÖTGLSKAASSIPSESAS-----KGLNÖMNN	2406	
Qy	2019	GHGNGÖGLTVVLOQÖGÖTTSQÖLO-----LIPÖGYTVLPGGÖQÖLMOAMNGTVÖRFL	2071	
Db	2407	GNCKANKVYELSRMSKSTSSGSESÖDRSERPVLVQSTFIEKASPTLRRKLESAS----	2461	
Qy	2072	FTPLATATATATTTTTVTSTAAGTÖQÖKSLSP-----QÖQVIOÖD-----KTLPPAQÖ	2120	
Db	2462	FESLSPSSRPSPSPRISQÖRPV-----LSPPLPMSLSTHSSVOAGWRRLPNNL	2511	
Qy	2121	SSSVGPKPAKPOPTÖQÖPÖFARPOPTÖQÖPÖPÖPÖVQÖPÖVQÖTVTVSSHVPSÖAQÖPTHAQÖ	2180	
Db	2512	SPTEIYWDGRPAKRHDLARHSHSESPPRLPINRGTMKRE---HSHKSSSLPVSSTWRTKG	2566	
Qy	2181	SSKPOVAASÖPOSNOVQÖSPRVAGSPÖTIRIPSTPSÖLS--PGÖQÖSÖYÖTTTÖPPIQÖ	2239	
Db	2569	SSSSILASSSESSEKASEDEKHAVNISGÖT--KÖSKENÖVAKSGTWIRKIKENEFSP----	2622	
Qy	2240	PHTSLOIPRQÖQPSQÖVQÖSSÖTÖTSSÖTGLNÖVSSVSPSRPÖLOIQÖPQÖVAYAPÖL	2299	
Db	2623	-----TNSISÖTVSSGÖT-----NGASKTLITÖMÄVAKSTEDV	2657	


```

Db 974 KRGOMKPSIESYSDHSDSCSYGQYPADLAHKIHSANHMDNDGDELDTPIINYSKAYSDE 1033
QY 552 --SOGESIRLSTKEVIMKGNINNYFRLGEGKYRYVHNOYSTNSFALNKHQREDHDK 608
Db 1034 QLNQSGRSPSONERWARPKHIEDEIKOSQROGR--NOSTT--YVYTESTDDKHLKE 1088
QY 609 RRLHAHFCLUPAGEFKWNSVHOSKVLITLRLITTOLENNIPSSFILPNNASHANW 666
Db 1089 QPHGQGEVSPYRSRANGS-----ETNKGVS-----NHGINQ 1122
QY 669 IKAQOMSKSRREFALALALECAVKPVYMLPIWREFLGHTRLMHMTIEREKEKVKKE 728
Db 1123 NVSGSLQOED-----DVEDDKPTVNSE 1144
QY 729 KQOEESTMOQAWKYTFPVKHQVWKQGEYRVGTGGMWISKTHVYRVPKLPDGT 788
Db 1145 RYSEEE-----QHEEERPTNY-----1161
QY 789 NVNYRKSLEGTKNMDEMDES-----DKRKSRSPPKKIKIEPDSEKD 831
Db 1162 SIKYNEE---KRVHDPIDISLKYATDIPSSOKOSFSPKSSGSSKTEHHMSSSENT 1217
QY 832 EVKSGDAKAGADQNMEDISKITEKKDQVVELDSDS-----DKP-CKEPEM 877
Db 1218 STPSNNAKR---QNLHPSSAQSRSGOPQKAATCKVSSINOETIQYCVEDTPICFGRCS 1274
QY 878 EYDDDMKTESIVNCOESSQ-----VDVYVNSGFIHRT-----SYKKTKSS 919
Db 1275 SLSSLSAEDIEICNOFTQOADSANTLOIAIEIKERIGTRSADPSEVPAVSOHPKTKSS 1334
QY 920 KLDLLEERRIKQFLEKORLEKIEKGIGIKGIGTSTNSKN-----LSESPVI--TKAK 973
Db 1335 RLQG-----SSLSSESARHKAVEFSSGAKSPKSGAQTPKSPPHYVQETPLMFBRCT 1387
QY 974 EGCOSDSMRQESPNANNDDPEDLIQCCSOSDSSVLRMSDPSTTNKLYPKRVLDDVSI 1033
Db 1388 SVSLSDSFESHASIVQSEF-----CSGAVSGIISPSPDIPDSCQTMPSR-----1434
QY 1034 RSPETKCPKONSJENDIEEKVSDLASRGOEPTKSKTGNGFFIDSKLASAD-----1085
Db 1435 --SKTPPPPOTAQTKRE-----VPKNKAPTAEKRESG-----PKQAAVNAVORVOV 1480
QY 1086 --DIGTLI-----CKNK-----KPLIOESDPIV-----SSSKSALHSV 1118
Db 1481 LPQADTLHLHATSTPDGFCSSSLALSLSDEPIQKDELRIIMPVQOENDNGETSEQ 1540
QY 1119 PKSTINDADAPILSRAMPEGLKCDSESNTLENSDPTVISIODSEEDMIVQNSNEISIE 1178
Db 1541 PKESNENO-----EKEAKTIDEKD--LLDSDDD-----1569
QY 1179 QFTRREDDVELEPLKCELVSGESTGNCEDRLPVKGTENGKRPQOKLEERPVNKCSD 1238
Db 1570 -----DIELE--EC-IISA-----MPTKSSR-KAKRAQAQASKLPPVYAKPS 1609
QY 1239 QI---KIKNTTKKNNNRESEKKGORTSTFOJINGKDNKPKIYLKGCLEISESRVSG 1295
Db 1610 QLPYKILPSPONRLQPKHVSFTPG-----DDMPRVY---CV-----EGTPINF 1650
QY 1296 NVEKVVNNINKIIPENDIKSLTVKESAIRPINDVIMEDFENENNSSETSKSHLSSDAE 1355
Db 1651 STATSLDLTIESPPNEL-----AAGEGVRCGAOSGEFEKDTIPTER--STDEAQ 1700
QY 1356 GNYRDSLETLPSTKES-----DSNQTTPPSASCP 1384
Db 1701 GGTKSSV-TIPELDNKAEBEDILAEICINAMPKGSHKPRPVKIMQVOQOASASSAP 1759
QY 1385 ESNSVNOVEDMEIETSEVKKVTs-----SPITSEESNLSN 1420
Db 1760 NKNQLDGKK--KKPTSPVKPIPOUTERYRVRKNNAKSKNNLNAERVPSDNKDSKQNLKN 1817
QY 1421 DFIDENGLPIKNKNVNGE-----SKRK 1443
Db 1818 NSKDFNKLPPNNEDEVGSPAFDPSHHYTPTEGTPYCFGRNDSLSLDFDDDDVLSREK 1877

QY 1444 TVITEVTTMTSVATESKTVIKVEKGDKQTVVSTENCASKSYT-----TTTTV 1493
Db 1878 ---AELKRAKENNESEKATSHTELTSNOSANKTQOIAQOPINRGKPILOKOSTPEQ 1934
QY 1494 TKLSTPSTGSDVIDISVKEOSKTVVTVTVTVVDSLTGTLVATSMVSKESYSTRDKVLM 1553
Db 1935 SKSDIPDRGAATD---EKLOMFALENTPVCFs---HSSLSLSLSDIDQENNNENEDIKE 1988
QY 1554 FSRP-----KTRSGTALPSYRK-----FYTKSTKSIIVLPNDL-----1589
Db 1989 TEPPDSQGEPSKPOASGVAPKSEFHEDTPVCFSSNSSLSSLSIDSEDDLQOECISSAMPK 2048
QY 1590 -KKIARKGIGIREVYFYNNKP-----ALDIWPPSPRPFGIT-----WRY 1630
Db 2049 KKRPSRLKQNE-----KSPRNMGILLEDLTLTDKIDQRPDSEHGLSDPENFDKMA 2102
QY 1631 RLQTVKSLAGVSLMLRLMLASLRMDMAKVPPOGSGSTRETSETTEITTEIIRKRDVGP 1690
Db 2103 IQEGANSIV-----SLHQAAAACL-----SRQASSDSD-----SILSKS---2139
QY 1691 YGIREYCIKKIICPIGVPEPTPKETPPQKKGLRSSALRPKRBEPTKQGPVILLETWAE 1750
Db 2140 -GI-----SLGSPFHLPDQOEKPTSNKGP--ILKPEKST-----LETKIE 2181
QY 1751 EE-----LELMEIRAFEREVEREKAQAVERQAKRLEQOKP-----TVIAT 1791
Db 2182 SESKIGIKGKKVYKSLTGTGVRNSF-----ISGOKQOPLOAMPISIKGRVITHI 2232
QY 1792 STTSPSTSTSTISPAOKVAVAPISGSVTTGTKNVLTGKVPATVTFPOONKNFHQTEAT 1851
Db 2233 PGVANSSTSTSPVSKKGPPLKTPASKSPSEGQVATTSPRAKPSV-----2277
QY 1852 WVKOGSGNSGVVOYQKVLGIIPBSTGTSOOTFTSPQPRATVYIRNTSGSGGTTNSQ 1911
Db 2278 ---KSELSPVARQTSQIG-----GSSK-----AFSRGSGRDSPTS--2309
QY 1912 VITGPQIRPGMTVIRTPLOOSTLGKALIRTPVMVOPG---APQVWTOIIRGQPVSTAV 1967
Db 2310 -----RPAQPLSRPIQSGRMS-----ISPRNGISPPNKLSQLPRTSSPSTA-2353
QY 1968 SAPNTVSTSPGQKSLTS-----ATSTNSIOSASQPPRPOGQVKTLMQLOLTQ 2018
Db 2354 ---STKSSGSGKMSYTPSGQMSQONLTQOTGLSKNASSIPRSESAS---KGLNOMN 2405
QY 2019 GHGNOGLTVYIOGQGTGLO-----LIPGVTVLPBGQGLQAMMPNGTVORFL 2071
Db 2406 GNGANKRVELSRMSSTKSSGESDRSERPVAVROSTIKAPRFTLRLKLEBAS-----2460
QY 2072 FTPLATTATTATTTTIVSTTAAGTGROKOSLSP-----QMOVHOD-----KTLPPAQ 2120
Db 2461 FESLSPSSRPASPRISQAOQIPV-----LSPSLPMSLSTHSSVQAGWRKLPPNL 2510
QY 2121 SSSVGPAAKAOPTQAPBARPOPTOPQSPAPQEPVOTOPPEVOTQTVSSHPSEKQPIHAQ 2180
Db 2511 SPRIEYNDGRAPRAKHDAHSSESPLRIPNRSQWRE---HSHKSSLSIPRVSTWRTG 2367
QY 2181 SSRPOVAOSQPOSNVQOGSPVPRVQSPQTRIRPSTRSOLS-POQSOVOVOTTSQPIPIQ 2239
Db 2568 SSSSILSASSESRKAKASEDEKHVNSISGT--KOSKENOVSAKGTWIKIKENEFSP----2621
QY 2240 PHTSLQIPSOQOPQPOPOVOSTOTLSSGQT 2270
Db 2622 -----TNSTSQTVSSGAT 2634

RESULT 12
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH

```


APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 2.4%; Score 337; DB 1; Length 2842;
Best Local Similarity 17.1%; Pred. No. 2.5e-11;
Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;

QY 387 EIRHNDITEDLTNRKAGSKSFLAA-ANEELIESIRAKKGDIDNYSKPETEKDKN---442
DB 801 DTRRHDDNNSDNTNT---GNMIVLSPLYLNTTVPSSSSSGSLDSRS---EKDKSLER 853
QY 443 -----ETENDSKDAEKNEEFEDOS-----LEKSDDKTDPDDPEOCK-----480
DB 854 ERGIGLVNYPATENPGTSSKRGLOJSTAAQOLAKYMEVSAIHTQGEDSSSTTELHC 913
QY 481 -----SEYGF-KSEKSN-----GELSESPGAG 502
DB 914 YVDERNALRRSSAAHTSHSTYNTFTKSENSNRCTSMFYAKLEYKRSSNDLSNVSDDGYG 973
QY 503 KGASGSTRITLRNDPSKLSQLKS--QOYAAAHAEANKL-FKEGKEVLVYN-----551
DB 974 KRQMKPSESIESDEDESKFCSTGYRPADIANKIHSANHHDDNDGELDPINISLKYSD 1033
QY 552 --SQGEISRLSTKKEYIMKGINNYFKLQGEKYYRYVHNOYSTNSFALNKHQREDHDK- 608
DB 1034 QLNSGCRSPSONRWRAPKHIIEDEIKOSQROSR--NQSTT--YPVYESTDDKHLKF 1088
QY 609 RRLIAKFCILTPAGEFKWNGSVGSKVLTITSLRLITTOLENNIPSSFLHPNASHRANW 668

DB 1089 QPHFGQDECVSPYRSKAGANS-----ETNRVGS-----NHCIQ 1122
QY 669 IKAOWMCKPREFALALALECAKVPVWLPIMREFLGHTRLHMSTIEREKYKKE 728
DB 1123 NVGSLQGED-----DVEDKPTVSE 1144
QY 729 KQDEETMOQATWVKTTPFKHQVWKQGEERYVGYGWSMISKTHYRFPVPLPGNT 768
DB 1145 RYSEEE-----QHDEEERPTVY-----1161
QY 789 NVNKRKSLGCTKNMDENMES-----DKRCSRPRKIKIEPOSEKD 831
DB 1162 SIKYNEE-----KRHVDPIDYSLKATYATDIPSSQKSEFSKSSGSSKTEHSSSENT 1217
QY 832 EVKSPDAKAGDONEMDISKITEKDDQVKELLDSOS-----DKP-CKEEPM 877
DB 1218 STPSNAKR--QNLHPSSAQSRSGQPOKAATCYKSSINQENTIQYCVBDFTICFRCS 1274
QY 878 EVDDMKTESHVNCQESSQ-----VDVNVSEGFHLRT-----SYKKTKSS 919
DB 1275 SLSSLSAEDEICGNQTOEADSANTLOIAIEIKETIGTSAEDPVSEVPASOHPRTKSS 1334
QY 920 KLOGLEERRIKQPLEKQKLEKIKLEGIKIGKGTNSKN-----LSESPVI-TKAK 973
DB 1335 RLOG-----SSLSSSAHHKAVESSGAKSPKSGAQTPKSPPEHYVOETPLMFSRCT 1387
QY 974 EGCQSDSMROGSPNANNQPEDLIQCSQSDSSVLRMSPPSHNTNKLYPKDRVLDSVI 1033
DB 1388 SVSSLDSFESRSIASSVQSEP-----CSGMVSGIISPLDPSPGQTMPPSR-----1434
QY 1034 RSPETKCPKONSIENDIEEKVSDLASRGQPTKSKTGNDFIIDSKLASAD-----1085
DB 1435 --SKTPPPPTQATKRE-----VPKNKAPTAEKRESG-----PKQAVNNAVQRYV 1480
QY 1086 --DIGTII-----CKNK-----KPLQESPTIV-----SSKSALHSSV 1118
DB 1481 LPADTLLHFATESTPDGFCSSSLALSLEDEPTQDVLRLIMPVQENDKNETESEQ 1540
QY 1119 PKSTNDRDATPLRAMDFEKKLGCDSSENSTLENSPTVSIQDSEEDMIVQNSNEISE 1178
DB 1541 PKSENEQ-----EKEAKTIDSEKD--LDDSDDD-----1569
QY 1179 QFRTREQDVEVLEPLKCELVSGESTGNCEDRLPVYKGTANGKRPQOKLEBRPVKCS 1238
DB 1570 -----DIEILE--EC-IISA-----MPTKSSR-KAKKPAQYASKLPPVARKPS 1609
QY 1239 QI---KIKNTDKKNNEERSEKKGQRTSTFOINGKDNKIKYLGKELKEISRYVSG 1295
DB 1610 QLEPVYKLLPQNRLOPKHVSFTPG-----DQMRVY---CV---EGTPIINP 1650
QY 1296 NVBPKVNNIKIIPENDIKSLTVKESAIRPFINGDVIMEDFNRNSETKSHLLSSDAE 1355
DB 1651 STATSLSDLTIESPPNCL-----AAGCVKGAOSGEFEKRDITPIEGR--STDEAQ 1700
QY 1356 GNYRDSLETLPTSKES-----DSTQTTTPSACP 1384
DB 1701 GKGTSVY-TIPELDNKAKEGDLIACINSAMPKSKHFRVYKIMDYQQAASASSAP 1759
QY 1385 ESNVNOVEDMEIETSEYKAVTS-----SPTTSEESLSN 1420
DB 1760 NKQLODKK--KKPTSPVKPIPONTEYRTVRNRNADSKNLNAERYFSNKKDKKONLKN 1817
QY 1421 DFIDENGLPINKNENNGE-----SKRK 1443
DB 1818 NSKDFNDKLPNNEBRVGRGSAFDSPHHYPIEGTPICFSNDSLSSLDPPDDVDLSREK 1877
QY 1444 TVTTEVYTMSTVATSEKTVIKVEKGDQTVVASTENCAKSTVY-----TTTIV 1493
DB 1878 --AELRKAKENKESPAKVYSHFELTNSQOASANKTQAIAPQINRQOPKPILOKQSTFPQ 1934
QY 1494 TKLSTPSTGSGVDITISVKEQSKTVVTTVYDLSLTGGTGLVYTMVYSKEYSTDKYKLMK 1553
DB 1935 SSKDIDPRGAATD--EKLQNFAIENTPVCF--HNSLSLSLSDIDQENNNKNEPIKE 1988

QY 729 KQOEETMQATWKTTPVKKHQQVKKQGEERYVYTGGMWISKTNYREVPLPGNT 788
 Db 1145 RYSEEE-----QHSEERPTNY----- 1161
 QY 789 NNRYRSLEGTKNNMNDENDES-----DKRCSRPKKIKIEPDSEKD 831
 Db 1162 STKYNE---RRHVDPIDYSLKYATPIDSSOKQSFSSKSSSGOSTEHSSSEENT 1217
 QY 832 EYKGSDAKADQONEMDISKITEKKQDQVKEKLLDS-----DKP-CKEEPW 877
 Db 1218 STPSSNAKR---QNOHPSSAQSRSQPKAACTCKYSSINQETIQTYCEDPTICSRCS 1274
 QY 878 EYDDDMKTESHVNOESSQ-----VDVNVNSEGFLHT-----SYKKTKSS 919
 Db 1275 STLSLSSADEICQNTQOADSANTLQIAIEIKETIGTSAEDPVEVPAVSQHPTKSS 1334
 QY 920 KLDGLLERIKQPTLEKORLEKIKLEGKIGKSTSSKN-----LSESVIT-TRAK 973
 Db 1335 RLOG-----SSLSESAHKAVERFSSGAKSPKSGAQTPKSPBHYVQETPLMESRCT 1387
 QY 974 ECGOSDMHQEOSPANNDQPEDLIGCCSDSSVLRMSDPSTTKLVPKRDVLDDVSI 1033
 Db 1388 SVSISLDSFESRSIASSVOSEP-----CSGWSGIISPDLPDPSQOTMPPSR----- 1434
 QY 1034 RSEPTKCPKQNSIENDIEKVSDDLASRGQEPPTKSTKGNDFIJDSEKLASAD----- 1085
 Db 1435 --SKTPPPPOQTAKRE-----VPKNAKPTAKKRESG-----PKQAVNAVQOVQY 1480
 QY 1086 --DIGLI-----CKNK-----KPLIOBESDTIV-----SSKSALHSSV 1118
 Db 1481 LPPADLLHFATESTPDGFCSSSLSLALSLDEPFIOKQDELRIIMPVOEDNQNETESQ 1540
 QY 1119 PKSTINDRATPLSRAMDFGKLCQDSESNSTLENSDPTVSIODSEEDMIVQNSNISE 1178
 Db 1541 PKESSNO-----EKEAKTIDSEKD-----LDDSDDD----- 1569
 QY 1179 QERTBQDVEVLEPLKCELVGSESTNCEDRLPVKTEANGKKPSQOKLEPPVKKCS 1238
 Db 1570 -----DIEILE--BC-IISA-----MPTKSSR-KAKKPAQTASKLPPVARKPS 1609
 QY 1239 QI---KLKMTQDKKNNESEKKGORTSTFOINGDNKPKIYLKCEKLEISESRVSG 1295
 Db 1610 QLEPVYKLLSPSONLOPKHVSFTPG-----DDMPRVY-----CV-----EGPPINE 1650
 QY 1296 NVEPKNNINKIIPENDIKSLFVKEBSAIPFINGDVIMEDFNRSSEKSHLLSSDAE 1355
 Db 1651 STATSLSDLTIESPPNEL-----AAGBVGAGQAGSEFEKRTIPIEBGR--STDEAO 1700
 QY 1356 GNYRDSLETLPTSTES-----DSTOTTTPSASCP 1384
 Db 1701 GCKTSSV-TIPELDONKABEGDILACINSAMPKGSHPFRVKKIMDQOQASASSAP 1759
 QY 1385 ESNSVNOVEDMELETSEVKKVIS-----SPTSEESNLSN 1420
 Db 1760 NKNQDLGKK--KKPISVPKPIQNTPEYRTVRKNADSKNNLNAERVPSDKOSKONLKN 1817
 QY 1421 DFIDENGLPINKKENVNGE-----SKRK 1443
 Db 1818 NKKDFDKLPNNEDRVKGSFAPDSPHHTPIEGTPLYCPSRNDLSSLDDEDDVDLSREK 1877
 QY 1444 TWITEVTMTSTVAVESKTVIKVEKQDQTVVSTENCARKSTVT-----TTTTTV 1493
 Db 1878 ---AELRAKAKENKSEBAKYTSHTELTSNQOASANKQAIKOPINRGOPKPILOKOSTFPQ 1934
 QY 1494 TKLSTPSTGSDVILSVKQOSTVYTTVTDSLTITGGGLVYMTYSKEYSTRDKYLK 1553
 Db 1935 SSKDIPDRGAAND---EKLONAIENTPVCFS--HNSSLSLSDIDQENNNKENEPIKE 1988
 QY 1554 FSRP-----KTRSGTALPSTYK-----FYTKSTKKSIFLPPNDL----- 1589
 Db 1989 TERPDSQGEPSKPAQAGYAPKSFHVEDTVPVCPSRNSSLSTLSDSEDDLQECISSAMRK 2048

QY 1590 -KLLARKGCIREFPYFNNAKP-----ADLIWPPSPRPFGIT-----WRY 1630
 Db 2049 KKKPSRLKQDNE-----KXSPRNMGGILGEDLTLDLKDIQRDSEHGLSPDSENDMKA 2102
 QY 1631 RLOQTVKSLAGVSLMLRLNASLWMDMAKVPKGGSGTRETSETETTTETTEIIRRDVGP 1690
 Db 2103 IOEGANSIV-----SSLHQAAAAACL-----SKOASSDSD-----SILSKS--- 2139
 QY 1691 YGIRFEYCIRKILICPIGVETPEKETPTPOKGLRSSALRRKREPTKQGPVLIETWAE 1750
 Db 2140 -GI-----SLGSPHLPDQGEKPTSNKGP--ILKRGKST-----LETKEIE 2181
 QY 1751 EF-----LELWETRAFAERVEKEKAQAVEQAKKLEQOKP-----TVIAT 1791
 Db 2182 SEKGKIGKGGKYYKSLITGVRSNSE-----ISGQMKOPLQANMPSISGRFMIIH 2232
 QY 1792 STTSPTSTSTSPAKQWVAVPISGSVYTGKMWLTJTKGSPATVTPQNNKHQOTPAT 1851
 Db 2233 PGVRNSSSTSPVSKGPPKTPAKSPSEGOYATTSPPKAPSV----- 2277
 QY 1852 WVKOGSNSGVVOQKVLGIIIPSTGTSQOFTSFQPRATVYTIPTNTSGSGGTTNSQ 1911
 Db 2278 -----KSELSPVARQTSQIG-----GSSK-----APSRGSRDSTPS-- 2309
 QY 1912 VITGPQIRPGMTYIRPLOOSTGLKAIIRTPVWVQG---APQVMTQIIIRGOPYSTAV 1967
 Db 2310 -----RPAQOPLSRPIQSPGRNS-----ISPGRMGISPPNKLSQLPRTSSPSTA- 2353
 QY 1968 SAPNTVSTPTGQKSLTS-----ATSTNSIQSSASQPPRRQOQOVKLTMAQLOLTQ 2018
 Db 2354 ---STKSSGSGKSYSPSGRQMSQMLTKQTLGSKNASTIPRESAS---KGLNOMN 2405
 QY 2019 GHGNOGLFVYIOGQOTTGOLQ-----LIPQGVTVLPQGOQLMAAMPNGTVQRF 2071
 Db 2406 GNAKANKVELSRKSSSTKSSSESDRSERPVLVQSTFIKAPPTLRKLEBAS----- 2460
 QY 2072 FYPPLATTAATSTTTTTVSTTAAAGTEQROSKLSP-----QMOVHOD-----KTLPPAO 2120
 Db 2461 FESLSRSSRAPSPTRSQAQTPV-----LSPSLPDKSLSTHSSVQAGWRKLPVNL 2510
 QY 2121 SSSVGPAAKQOPQAPQAPQOTOPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQ 2180
 Db 2511 SPPIEYNDGPRAPKRIHIAKSHSESPRLPIRSGTWKRE---HSHKSSISLPVSTWRRTG 2567
 QY 2181 SSKPQVAAQOSQOSNVQOSPVVQOSQOTRIRPSTPSQLS--PGQOSOVQTTISOPPIQ 2239
 Db 2568 SSSSILSASESESEKAKSEDEKHNISIGT--KQSENQVSAKGTARKIKENEFSP---- 2621
 QY 2240 PHTSLQIPSOQPOSOPOVQVSSQTOLSSQOT 2270
 Db 2622 -----TNSTSQTVSSGAT 2634

RESULT 14
 US-08-449-731-7
 ; Sequence 7, Application US/08449731
 ; Patent No. 6413727
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; ANAND, RAKESH
 ; CARLSON, MARY
 ; GRODEN, JOANNA
 ; HEDGE, PHILIP J.
 ; JOSLYN, GEOFF
 ; KINZLER, KENNETH
 ; MARKHAM, ALEXANDER F.
 ; MAKAMURA, YUSUKE
 ; THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 102
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,731
FILING DATE: 25-May-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/289,548
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7
Query Match 2.48; Score 337; DB 4; Length 2842;
Best Local Similarity 17.18; Pred. No. 2.5e-11;
Matches 384; Conservative 301; Mismatches 782; Indels 784; Gaps 93;
QY 387 EIRHMDITFDLTKKAGSKSLAA-ANELLLEIRAKKGDIDNVKSPETKDKN--- 442
DB 801 DTKRHDDNRSDNFT--GMYLSPYLNTVTLPLSSSSSKSLSSRS--EKDSLER 853
QY 443 -----ETENDSKDAEKNREFEEDOS-----LEKDSDEKTPDDPEOCK----- 480
DB 854 ERGIGLGNHPATENPGTSSKRGLQISTTAQAIKAMEEVSALHTSQEDRSSGTTLELC 913
QY 481 -----SEVGF-KSEKSN-----GELSESPGAG 502
DB 914 VDERNALRRSSAAHTHSNTYNTFTKSENSMPCMPYAKLEYKRSSNDLSNVSSSDYG 973
QY 503 KGASSGTRITTRLRNDKSLQSLKS--QVVAALAHANL-FREGKVLVYN----- 551
DB 974 KRGQMPSTLESSEDESKFCYGYRADLAHKIHSAHMDMDGDLTPINYSLSYSD 1033
QY 552 --SQGISRLSTKEVIMKGINNINNYFKLGOEGKRYVYHNOYSTNSPALNKHQHRDHDK- 608
DB 1034 QUNSGQSPSONRMAHPHIIIEDEIKOSEQORS--NOSTT--YPVYTESIDDKHLKF 1088
QY 609 RRLAHKFCILTPAGEFKMNGSVHGSVLTITSLRLITOLENNIPSSFLHPNVAHRANW 668
DB 1089 OHFPGQCECVSPYRSRGANS-----ETNRVGS-----NHGINQ 1122
QY 669 IKAIVGCSRPREFALALILCEAVKPVMLPTMREFLGHTRLHRMTSTREKEKAKKE 728
DB 1123 NVSOSLCQED-----DYEDDKPTNVSE 1144
QY 729 KROEEETMQATWVKYTFPVKHWKQKGEYRVYGGGWSMISKTHYRVFVKPLPGMT 788
DB 1145 KXSEEL-----QHEEERPTNY----- 1161

QY 789 NNAVRSLEGTKNMNMENDES-----DKRCSRPKKIKIEPSEKD 831
DB 1162 SIKYNEE-----KHHVQPIDISLAKYATDIPSSOKOSFSPKSSSGSSKTEHMSSENT 1217
QY 832 EVKSDAKGADQONEMDISKITEKKDQVKELLSDS-----DKP-CKEPM 877
DB 1218 STPSSNAKR--ONOLHPSAQRSGOPQKATCKVSSINQETIQYVEDPFCPSRCS 1274
QY 878 EVDMDKTEHSHVCOSSQ-----VDVNVSEGHRLT-----SYKKTKGS 919
DB 1275 SLSSLSAEDIEICNQTQOADSANTLQIAIEIKETGTSNAEDPVSEVPAVSQHPRTKS 1334
QY 920 KLDGLERRIKQFTLEKORLEKIKLEGKIGKSTNSKN-----LSESPVI-TRAK 973
DB 1335 RLOG-----SSLSSSAHKAVERFSSGAKSPSKGAQOTPKRPPHYQETPLMSRCT 1387
QY 974 EGCQSDSMROESPANNDPEDLIQCSQSDSSVLRMSDPSTTKLYPKRDVLDVSI 1033
DB 1388 SVSLDSFESRSIASSVOSEP-----CSQWVGIIISPDLPSQOTMPPSR----- 1434
QY 1034 RSPETCPKNSIENIEEKVSDLASRGQEPSTKNGDNFFIDSKIASAD----- 1085
DB 1435 --SKTPPPPTQATKRE-----VKNKAPYAEKRESG-----PKQAAVNAVORVOY 1480
QY 1086 --DIGTLI-----CKNK-----KPLIOESDTIV-----SSSKALHSSV 1118
DB 1481 LPDADTLHFAETSTPDGFCSSSLALSIDPEFIQDVLELRIMPVQENDNENEFESQ 1540
QY 1119 PKSTNDQATPLSRAMPBGLQCDSESNSTLENSSDTVSIQDSEEDMIVONSNEISE 1178
DB 1541 PKESNENO-----EKEAEKTIIDSEDD--LDDSDDD----- 1569
QY 1179 QFTRQDQVLEPLKCELVSGESTNCEDRLPVKGTBEANGKPKSQOKLEERPYNKCD 1238
DB 1570 -----DIEILE--EC-IISA-----MPTKSR-KAKKPAOTASKLPPVAKRS 1609
QY 1239 QI---KIKNTDKNNENRESEKKGQRTSTFQINGDKNPKIYLGECLEISESRVVG 1295
DB 1610 QLPVYKLPSONMLQOKHVSFPPG-----DDMPRYV--CV-----EGPIINF 1650
QY 1296 NVKPKVNNINKIIPENDISLYKESAIRPFIINGDIYMEDEFERNSEFKSHLSSDAE 1355
DB 1651 STATSLDLTIESPNEI-----AGEGVGAQSGEFKRDITPIEGR--STDEAQ 1700
QY 1356 GNYRDSLETLPTKES-----DSTQTTTPASCP 1384
DB 1701 GGTTSV-TIPELDDKAEGLIACINSAMPKGSHPFYKKTIMDYQOASASSAP 1759
QY 1385 ESNVNOVEDMEIETSEVKKVTs-----SPTSEESNLIN 1420
DB 1421 DPLDENGILINKNENNGE-----SKKK 1443
QY 1760 NKNQOLDGKR--KKTPSPVPIPONTEYRTVRKNADSKNNLAERYFSNKKSKONLKN 1817
DB 1418 NSKDFNDKLPNNEDRVKGFAPDSPHHYPIEGTPYCFSRNDSLSDLPDDDDVDLSREK 1877
QY 1444 TYTTEVTMTSVATSKVIVKVEKDKQTVVSTNCAKSTVT-----TTTTVV 1493
DB 1878 ---AELRAKKEKESPAKTSHTELTSNOQSAKQIQAIKQIINGQKRPILQOKSTPQ 1934
QY 1494 TKLSTPSTGSDVILSVKBQSKTVYTTVTDLSITGGTLVTSMTVSKYSTRDVKLWK 1553
DB 1935 SKMDIPRGAAND--EKJONFAIENTVPYCFs--HNSLSLSLSDIDQENNNKENEPIKE 1988
QY 1554 FSRP-----KTSNGTALPSYRK-----FVTSKKSITVLPNDL----- 1589
DB 1989 TEPPOSQEPSPKPOASQVAPKSFHVEDTPVCFSRNSSLSDIDEDDLQECISSAMPK 2048
QY 1590 -KILARKGIRREVYFENVAKP-----ALDIMYPSRPRTFGIT-----WRY 1630
DB 2049 KKKPSRLKGDNE-----KHSFRNMGILGEDTLTDLKDIOAPDSEHGLSPDSENFWDKA 2102


```

QY 926 ERIKOTLEKORLEKIKLEGKIGKIGTSTNSKN-----LSESPVI--TKAKEGQSD 979
Db 1340 -----SSLSSASARKHAKVEPPSGAKSPSKGAOTPKSPPHYVOETPLMSKRCSTVSSLD 1394
QY 980 SMROECPNANNDQPELDIOGCSQSDSSVLRMSDPHTTKLYPKDVLVDVSIKSPETK 1039
Db 1395 SFEERSJLASSVQSEPP-----CSGMSGIISPSDLPSDLPQOTMPPSR-----SKTP 1439
QY 1040 CPKONSIENDIEEKVSLASRGOEPTKSKTKGNDFFIDDSKLASAD-----DIGT 1089
Db 1440 PPPOTAOQTKRE-----VKNKAPTAEKRESG-----PKQAANVAVORVOYLPPADT 1487
QY 1090 LI-----CKNK-----KPLIOESDTIV-----SSSKALHSSVPKSTND 1124
Db 1488 LLIHATSETPPGFSCSSLSALSLDEFFIOKDVLRIMPVQENDNNGENESSEPKSNE 1547
QY 1125 RDATPLSRAMDEBGLGCDSESNSTLNSDPTVSIODSSEEDMIVQNSNISISEQFTRE 1184
Db 1548 NQ-----EKEAEKTIIDSEKD-----LLDDSDDD----- 1570
QY 1185 QDVEVLEPLKCELVSGESTGNCERLRYKGTENGAKKPSOOKKIEERPVKCSDOI--K 1241
Db 1571 -DIELLE--EC-IISA-----MPTKSSR--KGKPPAOTASKLPPPVAKRPSQLEPVYK 1616
QY 1242 LKNTTDKKNENRESEKKGQRTSTFOINGKDNKPKIYLKGECLKEISESRVSGNVBPKV 1301
Db 1617 LLPSQNRLOPKHVSFTPG-----DDMPRVY--CV-----EGPTINSTVATSL 1657
QY 1302 NNINKITIPENDIKSLTYKESAIRPPINGDVIMEDFENENSETSHLSSSDAAGNRDS 1361
Db 1658 SDLIIESPPNEL-----AAGEVNRGAQSGEFEKEDTIPTEGR--STDEAGGRTSS 1707
QY 1362 LETLPSTKES-----DSTQTTTPSACESNSVN 1390
Db 1708 V-TIPELDONKAEBSODLIAECINSAMPKSKHKKPPRYAKKIMDOYQOASASSAPKNOLD 1766
QY 1391 QVEDMEIETSEVKKYVS-----SPITSEESNLSDFIDEN 1426
Db 1767 GKK--KKPTSPVKPIPOUTEYRTVRYRKNADSKNNLNAERVPVSDMKSKQMLKNNSKDN 1824
QY 1427 GLPIKNENNVGE-----SKRTVITEV 1449
Db 1825 DKLPPNEDRVKGSFAFDSPHHYTPIEGTPYCFSRNDSLSDFDDDVDLSREK--AEL 1881
QY 1450 TMTSTVATESKTIVKVEGDKOFVVSSTENCASKTVT-----TTTTVTKLSTP 1499
Db 1882 RKAKNKSEAKVYSHTELTSNQOSANKTOALAKOPINRGOPKPILOKOSTFFPOSKDIP 1941
QY 1500 STGGSVDIISVKEOSKTVVTTVTDLSLTGTGTLVTSMTVSKKEYSTRDKVKLMKFSRP-- 1557
Db 1942 DRGAATD--EKLONFALIENTPVCS--HNSSLSLSDIDQENNNKENEPIKETEPDS 1995
QY 1558 -----KTKRSGTALPSYK-----FVTKTKKSTIFVLJPNDDL-----KKLAR 1594
Db 1996 OGEPSKPOASGAPKSFVEHEDTPVCFSRNSSLSDISEDDLOECISSAMPKKKPSR 2055
QY 1595 KGIREVYFNNNAKP-----ALDIWYPSPRPTFGIT-----WRYRLQTVK 1636
Db 2056 LKGDME-----KHSPRMMGILGEDLTLDKIDQRPDSHGLSPDSSENFDMKALQEGAN 2109
QY 1637 SLAGVSLMLRLIMASLRMDDMAKVPPOGGGSTRFETSETETTTETITIKRDVGYCIKFE 1696
Db 2110 STV-----SSIHQAAAAACL-----SROASSDSD--SILSLKS--GI-- 2142
QY 1697 YCIRKIICPIGVETPKETPTPQORKGLSSALRPKRPTPKOTGPVILLETWAAEEE-- 1752
Db 2143 -----SLGSPFHLTPDOEEKPPTSNGPR--IKKPEKST-----LETKKISESKGI 2188
QY 1753 -----LELMELIRAFEREREKEAOAVEQOAKKRLQOKP-----TVIATSTTSP 1797
Db 2189 KGGKRVYKSLITGKVRNSN-----ISGQMKOPILOANMPSLSRGRTHIHIGVANS 2239
QY 1798 SSTSTISPAQKVVAPISGVSVTGTKNVLTTKVGSFATVTFQONKNFHOTFAIWKOGO 1857

```

```

Db 2240 SSTSPVSKKGPPLKLPKAPSKSPSEGOATATSPRGAKPSV-----K 2279
QY 1858 SNGGVVOVQOKVIGITIPSTSTGTSQOTFTSFQORTAVITIRPMTSGSGGTTNSQVITGPQ 1917
Db 2280 SELSPVARTSOAQ-----GSSK-----APSRSGRSDTSP----- 2310
QY 1918 IRPGMTVIRTPLOOSTLGAIRITPVMAVPG-----APQOVMTQIIRGQFVSTAVSAPNTV 1973
Db 2311 -RAAQOPLSRPIOSPGRNS-----ISGRNGISIPPKKLSQLPRTSSPSTA--STK 2357
QY 1974 SSTPGKSLTS-----ATSTNISQASQPPRPQOGQVKLYMAQULTQLTQHGQNO 2024
Db 2358 SSGSGKMSYSPGRQMSQOONLTKQTGLSKNASSIPRSESAS-----KGLNOMNNGANK 2412
QY 2025 GLTVIVUGGQOTTGQLO-----LIPQVTVLPGGQQLMQAAMPNGTVQFPLFTPLAT 2077
Db 2413 KVELSRMSSTKSSGSESDRSERPVLVROSTFIEKAFSPTLRRLKLESAS-----FESLSP 2467
QY 2078 TATTAATTTTTSVSTAAGTGEOROSKLSP-----QMQVHOD-----KTLPPAQSSSVGP 2126
Db 2468 SSRPASPTRSOAQTPV-----LSPSLPDMSLSTHSSVQAGGWKRLPMLSPTIEX 2517
QY 2127 AKAOPTOAPSARPQOTOPQSPAQPEVOTQPEVOTQTVSSHVPSQAPTHAOSKRPV 2186
Db 2518 NDRGRPAKRHDIAKSHSESPPRLPINRSGTWKRE--HSKHSSLPVSVTWRTGSSSSTL 2574
QY 2187 AOSQOPSANVOGSPVAVOSPQTRIRPSPQOLS--PGOOSQVQTTTSQPIPIOPHTSLQ 2245
Db 2575 SASSESEKAKSEDEKHVNSISGT--KQSKENQVSAKGTWRIKENEFSP----- 2622
QY 2246 IPSQOGPOSOPOVQOSTOTLSSGOT 2270
Db 2623 -----TNSTSQTVSSGAT 2635

```

Search completed: November 20, 2002, 16:34:16
 Job time: 142.024 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:01 ; Search time 99.1487 Seconds
(without alignments)
6041.216 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971
Sequence: 1 MVSEEEEDGDGAETQDSE.....KLGFKASRSHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvivirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14170	94.6	2781	4 09UG2	09UG2 homo sapien
2	2771.5	18.5	2669	5 09SVB8	09SVB8 drosophila
3	2705.5	18.1	645	11 08VDN7	08VDN7 mus muscula
4	1488.5	9.9	976	5 09W0T0	09W0T0 drosophila
5	1368.5	9.1	412	4 09H5E0	09H5E0 homo sapien
6	854.5	5.7	1711	5 04S409	04S409 caenorhabdi
7	848.5	5.7	1713	5 09S5U8	09S5U8 caenorhabdi
8	692.5	4.6	1022	5 0960Y3	0960Y3 drosophila
9	661.5	4.4	997	5 09W0T1	09W0T1 drosophila
10	621.5	4.2	2768	5 09VC00	09VC00 drosophila
11	613.5	4.1	5476	5 09AV17	09AV17 drosophila
12	613.5	4.1	5533	5 09VPL2	09VPL2 drosophila
13	613.5	4.1	5554	5 09VPL1	09VPL1 drosophila
14	613.5	4.1	5560	5 09VPL1	09VPL1 drosophila
15	608.5	4.1	5533	5 09UC63	09UC63 drosophila
16	593	4.0	17352	5 095YM2	095YM2 procamburus

17	571.5	3.8	16215	5 09NFS3	09NFS3 drosophila
18	567	3.8	6815	5 0917U4	0917U4 drosophila
19	562.5	3.8	3257	5 09V736	09V736 drosophila
20	561	3.7	3080	5 09VRY3	09VRY3 drosophila
21	551.5	3.7	2897	5 09VTD9	09VTD9 drosophila
22	546.5	3.7	5327	5 076891	076891 drosophila
23	518.5	3.5	2081	10 09LH98	09LH98 arabidopsis
24	513.5	3.4	2465	5 022463	022463 caenorhabdi
25	508	3.4	2910	10 09FND5	09FND5 arabidopsis
26	503.5	3.4	4498	5 09W223	09W223 drosophila
27	496	3.3	6994	5 017343	017343 caenorhabdi
28	490	3.3	5170	5 017490	017490 caenorhabdi
29	489	3.3	3420	2 09AER7	09AER7 staphylococ
30	474	3.2	3726	5 09VFL1	09VFL1 drosophila
31	463.5	3.1	3507	5 023587	023587 caenorhabdi
32	462	3.1	2586	5 09VTK8	09VTK8 drosophila
33	460.5	3.1	3484	5 P91257	P91257 caenorhabdi
34	456.5	3.0	3261	4 09Y556	09Y556 homo sapien
35	456.5	3.0	3664	4 096Y58	096Y58 homo sapien
36	453	3.0	3551	11 099PS4	099PS4 mus musculu
37	451	3.0	5085	11 09JKS6	09JKS6 rattus norv
38	449.5	3.0	3111	5 09VH10	09VH10 drosophila
39	449	3.0	2283	2 08V099	08V099 staphylococ
40	444	3.0	2977	5 09VAP9	09VAP9 drosophila
41	441.5	2.9	2083	5 09N435	09N435 caenorhabdi
42	441.5	2.9	5120	13 09PU36	09PU36 gallus gall
43	440.5	2.9	3576	11 09Q2W2	09Q2W2 mus musculu
44	440	2.9	2271	16 0980Y4	0980Y4 staphylococ
45	440	2.9	2471	13 09DDN8	09DDN8 xenopus lae

ALIGNMENTS

RESULT 1

09UG2 PRELIMINARY; PRT; 2781 AA.

AC 09UG2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Bromodomain PHD finger transcription factor.
GN BPTF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20130111; PubMed-10662542;
RA Jones M.H., Hamana N., Shlman M.;
RT "Identification and characterization BPTF, a novel bromodomain
transcription factor."
RL Genomics 63:35-39(2000).
DR EMBL; AB032251; BAA89208.1; -.
DR HSSP; O92831; 1B91.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT_dom.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
SQ SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1AC3CFBE CRC64;

Query Match 94.6%; Score 14170; DB 4; Length 2781;
Best Local Similarity 95.1%; Pred. No. 0;

Matches 2766: Conservative 2; Mismatches 13; Indels 126; Gaps 1;	
QY 1 MYSEEEEDDADAEETODSEDEDEDEEDDDSDYPEEMDDDDDAASYCESSFRSHST 60	
Db 1 MYSEEEEDDADAEETODSEDEDEDEEDDDSDYPEEMDDDDDAASYCESSFRSHST 60	
QY 61 YSSTPGRRKPVHHRSPRIIEEKDIPLEFPKSSSEDLMPNEHIMNYIAIEVLRNFQTV 120	
Db 61 YSSTPGRRKPVHHRSPRIIEEKDIPLEFPKSSSEDLMPNEHIMNYIAIEVLRNFQTV 120	
QY 121 LRLSPFRFEDCALVSOEOTLMAEMHYLLKAVLRBEDTSNTTGPADLKDSVNSTLY 180	
Db 121 LRLSPFRFEDCALVSOEOTLMAEMHYLLKAVLRBEDTSNTTGPADLKDSVNSTLY 180	
QY 181 FIDGTMPEVLRVYCESSEKEXHNLPLYOEADYPRGPENKIKYLOFIVDOFLTNTIARE 240	
Db 181 FIDGTMPEVLRVYCESSEKEXHNLPLYOEADYPRGPENKIKYLOFIVDOFLTNTIARE 240	
QY 241 ELMSEGVIOYDHCRCVKRLDGLCCETCSAVYHLECVKPPLEEVPEDEMOCEVCAHKV 300	
Db 241 ELMSEGVIOYDHCRCVKRLDGLCCETCSAVYHLECVKPPLEEVPEDEMOCEVCAHKV 300	
QY 301 PGVYDCAVIEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIEEDTENEKEKIWTYSTVQ 360	
Db 301 PGVYDCAVIEIOKNKPYIRHEPIGYDRSRKRYWFLNRRLIEEDTENEKEKIWTYSTVQ 360	
QY 361 LAELIDCDKDYWEALCKILEEMREEIHRHMDITEDLTNRKARSNKSFLLAAMEILIES 420	
Db 361 LAELIDCDKDYWEALCKILEEMREEIHRHMDITEDLTNRKARSNKSFLLAAMEILIES 420	
QY 421 IRAKGDIDNYSPEETEKDNETENDSKDAEKNREFEEDOSLEKSDDKTPDDDEPGOK 480	
Db 421 IRAKGDIDNYSPEETEKDNETENDSKDAEKNREFEEDOSLEKSDDKTPDDDEPGOK 480	
QY 481 SEEPTEVODKNSVANSNGDVTNATSEETSPSEGRSPVGLSETPDPSNMAEKKVASL 540	
Db 481 SE----- 482	
QY 541 PQDYPEEPNKTCESSNTSATTTTSIQPLNLENSNSSSELNSSQSESAKADDENGERESHST 600	
Db 483 ----- 482	
QY 601 PVSIOEEIVGDFTSKSTGELSESPPGAKGASGSTRITTRLRNPDSKISQIKSQOYVAAA 660	
Db 483 -----VGFKESEKSENGELSESPPGAKGASGSTRITTRLRNPDSKISQIKSQOYVAAA 534	
QY 661 HEANKLFKEGKEVLYVNSOGELSRISTKEEVIAMKGINNRYKLOGEYRYVHNOYSTNS 720	
Db 535 HEANKLFKEGKEVLYVNSOGELSRISTKEEVIAMKGINNRYKLOGEYRYVHNOYSTNS 594	
QY 721 FALNKHQREDHDKRRHLAHRFCLTLPAGEFKWNSVHSGSVYLTSTLRLLTTQLENNIPS 780	
Db 595 FALNKHQREDHDKRRHLAHRFCLTLPAGEFKWNSVHSGSVYLTSTLRLLTTQLENNIPS 654	
QY 781 SFPHNMAASHRANMTKAYONGSKPREPALALALIECAVKPVMPLPIWBEFLGHRRLHMT 840	
Db 655 SFPHNMAASHRANMTKAYONGSKPREPALALALIECAVKPVMPLPIWBEFLGHRRLHMT 714	
QY 841 STEREKEXVKKKEKOESEETMOATWVYTFPVKHQVMKOGEEYVVTGCGGSMWISK 900	
Db 715 STEREKEXVKKKEKOESEETMOATWVYTFPVKHQVMKOGEEYVVTGCGGSMWISK 774	
QY 901 THVRFVFKLPGNTVNVNRKSLGRTKNMDEMDESDRKCSRSFPIKIEPDSKEDVK 960	
Db 775 THVRFVFKLPGNTVNVNRKSLGRTKNMDEMDESDRKCSRSFPIKIEPDSKEDVK 834	
QY 961 GSDAKGADONEMDISKITEKKODVFKLLDSDDPKCKEEMPEVDDDMKTESHVNCOS 1020	
Db 835 GSDAKGADONEMDISKITEKKODVFKLLDSDDPKCKEEMPEVDDDMKTESHVNCOS 894	
QY 1021 SOVDVNVNVEGFHLRTSVKTKKTKSSKLDGLERRIKOFTLEKOKLEKIKLEGKIKGIGK 1080	
Db 895 SOVDVNVNVEGFHLRTSVKTKKTKSSKLDGLERRIKOFTLEKOKLEKIKLEGKIKGIGK 954	

QY 1081 TSTNSKNLSESPYITKAKGCSDSMRQOSPANNDDQEDLIQCSQSQSDSVLRMSDP 1140	
Db 955 TSTNSKNLSESPYITKAKGCSDSMRQOSPANNDDQEDLIQCSQSQSDSVLRMSDP 1014	
QY 1141 SHTTNKLYPKDRVLDDVSIRSPEFTCKPONSIEDEIEKYSDLASRQOETKSKTKGND 1200	
Db 1015 SHTTNKLYPKDRVLDDVSIRSPEFTCKPONSIEDEIEKYSDLASRQOETKSKTKGND 1074	
QY 1201 FIDSKLASADIGTLCKKKKPLIOESDPTIYSSKSALHSSVPKSTNDRDAPTLSRAM 1260	
Db 1075 FIDSKLASADIGTLCKKKKPLIOESDPTIYSSKSALHSSVPKSTNDRDAPTLSRAM 1134	
QY 1261 DFECKLGDSESNSTLENSDPTVSIQDSSEDMIVQNSNESISEQFTREQDVEVLEPLK 1320	
Db 1135 DFECKLGDSESNSTLENSDPTVSIQDSSEDMIVQNSNESISEQFTREQDVEVLEPLK 1194	
QY 1321 CELVSGSTGNCEDRLPVKGTENGAKRPQOQKLEERPVKCSQOIKLKTMTTKKNENR 1380	
Db 1195 CELVSGSTGNCEDRLPVKGTENGAKRPQOQKLEERPVKCSQOIKLKTMTTKKNENR 1254	
QY 1381 ESEKKGQRTSFQINGKDNKPKIYLKGECLKEISESRVSGNVEPKVNNINKIIPENDIK 1440	
Db 1255 ESEKKGQRTSFQINGKDNKPKIYLKGECLKEISESRVSGNVEPKVNNINKIIPENDIK 1314	
QY 1441 SLTVKESAIRPFINGDVIEMEDFENRNSSETRKSHLLSSDAEGNYSRSLFTLPSTKESDST 1500	
Db 1315 SLTVKESAIRPFINGDVIEMEDFENRNSSETRKSHLLSSDAEGNYSRSLFTLPSTKESDST 1374	
QY 1501 OTTPPSACPESNVNOVEDEMETSEVKKYTSPTISEESNLSNDFIDENGPIPKNE 1560	
Db 1375 OTTPPSACPESNVNOVEDEMETSEVKKYTSPTISEESNLSNDFIDENGPIPKNE 1434	
QY 1561 NVNGESKRKYTYTEVTMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVT 1620	
Db 1435 NVNGESKRKYTYTEVTMTSTVATESKTVIKVEGDKQTVVSSSTENCAKSTVTTTTVT 1494	
QY 1621 KLSPTSTGGSVYDIISVKEQSKTVVTTVTDSLTGTLVTSMYSEKYSTRODKVLMKF 1680	
Db 1495 KLSPTSTGGSVYDIISVKEQSKTVVTTVTDSLTGTLVTSMYSEKYSTRODKVLMKF 1554	
QY 1681 SRPKTRSGTALPSYRKFTVNTSTKKSIFVLPNDLKLARKGIREVPYNYNAKPALDI 1740	
Db 1555 SRPKTRSGTALPSYRKFTVNTSTKKSIFVLPNDLKLARKGIREVPYNYNAKPALDI 1614	
QY 1741 WPPSPREFTGITWRKRYLRQTVKSLAGVSLMLRLMASLRMDMAKVPFGGSTRFETSE 1800	
Db 1615 WPPSPREFTGITWRKRYLRQTVKSLAGVSLMLRLMASLRMDMAKVPFGGSTRFETSE 1674	
QY 1801 TEITTELIRKRDVGPPIGIREYCIKRICTIGVPEPKETPTPQKGLASSALRPKRE 1860	
Db 1675 TEITTELIRKRDVGPPIGIREYCIKRICTIGVPEPKETPTPQKGLASSALRPKRE 1734	
QY 1861 TPKOTGPVIEETWVAEELELMEIRAFERERKKAQAVEQAKRLLEQOKPVIATST 1920	
Db 1735 TPKOTGPVIEETWVAEELELMEIRAFERERKKAQAVEQAKRLLEQOKPVIATST 1794	
QY 1921 SPTSSTSTISPAQKVMVAPISGSVTTGKTAVLTTKVGSFATVTFQONKNFHQTEATVWK 1980	
Db 1795 SPTSSTSTISPAQKVMVAPISGSVTTGKTAVLTTKVGSFATVTFQONKNFHQTEATVWK 1854	
QY 1981 QGOSNGVVOYQKAVLGIIPBSTGTSOOTFTSFOPRATVYIRANTSGSGSTTNSQVIT 2040	
Db 1855 QGOSNGVVOYQKAVLGIIPBSTGTSOOTFTSFOPRATVYIRANTSGSGSTTNSQVIT 1914	
QY 2041 GPOIRPGMTVIRTPLOOSTGLKATIRTPVNVVOPAPQOVMQOIRGQPVSTAVASAPTVS 2100	
Db 1915 GPOIRPGMTVIRTPLOOSTGLKATIRTPVNVVOPAPQOVMQOIRGQPVSTAVASAPTVS 1974	
QY 2101 STPOKSLTASATSTNSIQSSASQPPRPOQGVKLTMAQLTOLQGHGNGSLVTVIOGOG 2160	
Db 1975 STPOKSLTASATSTNSIQSSASQPPRPOQGVKLTMAQLTOLQGHGNGSLVTVIOGOG 2034	

QY	2161	QTGQLDLIPGVTVLPBGQQLMOAMAMPNGVYQRLPFLPATYTTATSTTTTSTVSTAA	2220
Db	2035	QTGQLDLIPGVTVLPBGQQLMOAMAMPNGVYQRLPFLPATYTTATSTTTTSTVSTAA	2094
QY	2221	GTGGOROSKLSPMQVNOVDLTPLPAAOSSSGAPKAAPOTPAORASAPROTOPOQAPQEV	2280
Db	2095	GTGGOROSKLSPMQVNOVDLTPLPAAOSSSGAPKAAPOTPAORASAPROTOPOQAPQEV	2154
QY	2281	QTOPEVOYOTTTVSSHVSEAOPTTHAASSKPKOVAASQOPSNVOGOSPVRVOSPQTRLR	2340
Db	2155	QTOPEVOYOTTTVSSHVSEAOPTTHAASSKPKOVAASQOPSNVOGOSPVRVOSPQTRLR	2214
QY	2341	STPSQLSPGQSOYQTTTSPPTIPQHTSLQIPSGQPOSPQPOVOSTQTLSSGOTLNV	2400
Db	2215	STPSQLSPGQSOYQTTTSPPTIPQHTSLQIPSGQPOSPQPOVOSTQTLSSGOTLNV	2274
QY	2401	SVSSPSRPOLOIOOPQVAVAPBOLQOQVQVLSQISOQVAVQLOAQOGSGVPOQIKLOPI	2460
Db	2275	SVSSPSRPOLOIOOPQVAVAPBOLQOQVQVLSQISOQVAVQLOAQOGSGVPOQIKLOPI	2334
QY	2461	QIOQSAVOTHTQLONVTVQVAASVOBOLRVQQLRQOQKKRQOQIETINVTPSKLQI	2520
Db	2335	QIOQSAVOTHTQLONVTVQVAASVOBOLRVQQLRQOQKKRQOQIETIKREHTLOASQNS	2394
QY	2521	EIIQOVYMKHNAVIEHLKOKKSMTPAEREENQMTVCNQVKKYILDKIDKEKOAAKR	2580
Db	2395	EIIQOVYMKHNAVIEHLKOKKSMTPAEREENQMTVCNQVKKYILDKIDKEKOAAKR	2454
QY	2581	KRESVEOKRSKQAKTSLALFKHKEOURLFKRKRALDDLOLEVOEELKRLKIK	2640
Db	2455	KRESVEOKRSKQAKTSLALFKHKEOURLFKRKRALDDLOLEVOEELKRLKIK	2514
QY	2641	EKDLMOLOAQATVAVAPCPPTVPLPAPAPAPSPPPPPGVQHTGLSTPTLPVASOKRR	2700
Db	2515	EKDLMOLOAQATVAVAPCPPTVPLPAPAPAPSPPPPPGVQHTGLSTPTLPVASOKRR	2574
QY	2701	EEBKDSSSKSKKKMISTSTSEKFKKDTKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI	2760
Db	2575	EEBKDSSSKSKKKMISTSTSEKFKKDTKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI	2634
QY	2761	LOSELELIDEVQCCOSTEDAMTVTLPTLEKDYBGLKRVLSLOAHKKNAPFLEVPDN	2820
Db	2635	LOSELELIDEVQCCOSTEDAMTVTLPTLEKDYBGLKRVLSLOAHKKNAPFLEVPDN	2694
QY	2821	DAPDYGVIKEMPDLAMEERVOARRYEKLTFEVADMTKIFPNCRYVNSPSPFOCAEV	2880
Db	2695	DAPDYGVIKEMPDLAMEERVOARRYEKLTFEVADMTKIFPNCRYVNSPSPFOCAEV	2754
QY	2881	LESFFVOKLKGFKASRSHNNKLOSTAS	2907
Db	2755	LESFFVOKLKGFKASRSHNNKLOSTAS	2781
RESULT 2			
ID	Q95VB8	PRELIMINARY:	PRT: 2669 AA.
AC	Q95VB8		
DT	01-DEC-2001 (Tremblrel. 19, Created)		
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Nucleosome remodeling factor large subunit NURF301.		
GN	EFBX) OR NURF301 OR CG7022 OR CG1135.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:		
CC	Ephyridioidea, Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha:		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21468388; PubMed=11583616;		
RA	Xiao H., Sandalitzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,		
RT	Fu D., Wu C.;		
Dual functions of largest nurf subunit nurf301 in nucleosome sliding			

[illegible]


```
Db 2353 -LEHHELLAKRSLRLSLERMLQSEIHE---DVKTQVRHVRPLSNA----- 2397
Qy 2661 TPVLPAPAPAPSPPPPPQVQHTGLSTPLPY-----ASQKRKREEK--- 2704
Db 2398 -----SPDQSENERGSG---EPRLDPKRTQVONPRHGAQRPKLTTRKKRKYLC 2442
Qy 2705 -----DSSSKSKKKMISTTSKTRKTDTRKLYCICK 2734
Db 2443 ICTRPYDTRKRYVYVCDLCSNMFFHGDVCSITEEASK-KLSEFICIDCKRAETQOLYCSQR 2501
Qy 2735 TPVDESKFYICDRCQNMWYHGRVGLQSEAEILIDEVYCPQCCSTEDAMTV-LTPILEKD 2793
Db 2502 QPDESQFETICCDKQCDWFGHRCVGIQSEAEFIDEVYCPCECQKNDANANMKKLTLSND 2561
Qy 2794 YELIKRVLRLSLQAHKMAKPFLEVPDNDAPDYGYIKRPMDLATMEERVQRYREKLTTER 2853
Db 2562 VEBLTKMLINQOMLHKSAMFEMERVPDKEADYKYVKEKPMDLKRMELKLESNTYTKLSER 2621
Qy 2854 VADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVOKLKGRF 2893
Db 2622 IGDMTKIFDNCRYNPKESSEFFYKCAEALSESFFYOKIKNFR 2661
```

RESULT 3

```
Q8VDN7 PRELIMINARY: PRT: 645 AA.
ID Q8VDN7
AC Q8VDN7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 72.2 kDa protein (Fragment).
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021489; AAH21489.1; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001965; EGF-like.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00628; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 645 AA: 72167 MW: 6486F34F339820F9 CRC64:
```

Query Match 18.1%; Score 2705.5; DB 11; Length 645;
Best Local Similarity 82.9%; Pred. No. 1.2e-100;
Matches 548; Conservative 33; Mismatches 55; Indels 25; Gaps 10;

```
Qy 2251 GKAQAQAPQASARPOPTQPOSAPQAPQVQTPQVQOTVSSHWPSKQAPHAOSKRP 2310
Db 4 GAQAAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQ 55
Qy 2311 GYAQAQSPQSNVQSGSPFVAVQSPQTRIRPSPQSLSPGQSGVQVQTTQSPRIQPTSL 2370
Db 56 LVATQCPQSSSVQSGSPFVAVQSPQTRIRPSPQVQTPQAPQAPQAPQAPQAPQAPQ 115
Qy 2371 QIPSGQSPQSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ 2430
Db 116 QAPSGQSPQSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQ 169
Qy 2431 VLSQIOSVVAQIQAOQSGVPOQIKQLPFIQIQOQSSAVQTHQIQNVVTVQASVQELOR 2490
```

```
Db 170 VLSQIOSVVAQIQAOQSGVPOQIKQLPFIQIQOQSSAVQTHQIQNVVTVQASVQELOR 226
Qy 2491 VQOLROOQOKKKQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 2550
Db 227 VQOLROOQOKKK-QQIETEHEHTLQASNOSEETIQOQVAKHNAVIEHLQOKKTMTAERE 285
Qy 2551 ENQRMIVCNQVMKYLLDKIDKEKQAKKRRKESEVQKRSKQONATRLSLFKRKEQRL 2610
Db 286 ENQRMIVCNQVMKYLLDKIDKEKQAKKRRKESEVQKRSKQONATRLSLFKRKEQRL 345
Qy 2611 AELIKRALLDKQLQIYQVQELKRDIKIKKEDLMQLAQTAVAACPPTPLPAP-PA 2669
Db 346 AELIKRALLDKQLQIYQVQELKRDIKIKKEDLMQLAQTAVAACPPTPLPAPAPAPA 405
Qy 2670 PPPSP-PPPGVQHGLSTPT-LPVASQKRRKESEVQKRSKQONATRLSLFKRKEQRL 2725
Db 406 APPAPRSPSPSTHSLPPAGHPAPPLPVTSQKKRKEEDQ-SKSKKKMISTTSKARK 463
Qy 2726 DTKLYCICKTPYDESKFYICDRCQNMWYHGRVGLQSEAEILIDEVYCPQCCSTEDAMTV 2785
Db 464 DTKLYCICKTPYDESKFYICDRCQNMWYHGRVGLQSEAEILIDEVYCPQCCSTEDAMTV 523
Qy 2786 LTPLETKDYEGKRLVRLSLQAHKMAKPFLEVPDNDAPDYGYIKRPMDLATMEERVQRR 2845
Db 524 LTPLETKDYEGKRLVRLSLQAHKMAKPFLEVPDNDAPDYGYIKRPMDLATMEERIKOR 583
Qy 2846 YYEKLTPEVADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVOKLKGRFKASRSNNKLOST 2905
Db 584 YYEKLTPEVADMTKIFDNCRYNPSDSPPFYQCAEVLSEFFVOKLKGRFKASRSNNKLOST 643
Qy 2906 A 2906
Db 644 A 644
```

RESULT 4

```
Q9WOTO PRELIMINARY: PRT: 976 AA.
ID Q9WOTO
AC Q9WOTO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CG17135 protein (Fragment).
GN E(BX) OR CG1022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Neilson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
```


Db	943	-----	KKRGRS-	-----	GKS	952
QY	1082	STNSKNI,SESPVITAKAKEGQSDSMFOEDSPMANNDQPEDLIQGSQSDSVLRMSDPS	114			
Db	953	SKPPFEFLT-----	NGCSPSCRS-	-----	971	
QY	1142	HTTKNLYKDAVLDDVSI,RSFPTPCPRQNSIENDIERKVSDLASRQDEPTKXTKGNDF	120			
Db	972	-----	-NNRKC-	-----	980	
QY	1202	IDDSKLASADDIGTLICKNNKPKLLQIESDPTIVSSSKSALHSVPKSTNDRDAPLSRAMD	126			
Db	981	-----	-MGRN-	-----	984	
QY	1262	FEGLICDSESNSTLENSSDTVSIQDSEEDMIIVQNSNESISEQFTRBODEVLEPLKC	1322			
Db	985	-----	-----	-----	984	
QY	1332	ELVSGSEPTGNEDL,PKYGTFRANKRSQOKKLEERPVNKCSDQIKLTKTTDKNNENE	138			
Db	985	-----	GLVSAKQAHDRKLEE	-----	1001	
QY	1382	SEKKQRTSTFOJGKDNKPKRIVLKGCELKEISESRVSGNVEPKVNNINKIIPENDIKS	144			
Db	1002	-----	-----	-----	1001	
QY	1442	LTVKESAIRPFINGDVI,MEDFNERNSSSETKSHLLSSSDAGNYRDSL,ETLPLSTKESDYO	1507			
Db	1002	-----	-----	-----	1001	
QY	1502	TTTTASCPESNSNOVEDMEIETSEPKKATYTSPITSEBSNLSNDFIDENGLPIPKNEN	156			
Db	1002	-----	-----	-----	1003	
QY	1562	VNGESKRKTIVTEVTMTSTVATESKTIV,KEGDKQTVWSTPENCASKTVTTTTYTK	1621			
Db	1004	VLGEK	-----	-----	1005	
QY	1622	LSTPSTGASVDI,ISVKEOSKTVVTTVTDS,LLTTGGTLVTSMTVSKEYSTRDKVLMKFS	1681			
Db	1010	-----	-----	-----	1005	
QY	1682	RPKTRGSTAL,PSYRKVTSTYKSTKSTFV,LPNDLKLAKRGGREY-PIYFNMAKPALD	1733			
Db	1010	-----	AMP,PELOFTSTRGKRS,IFLOKKILROMIMGGOQOYYMFGSAGIKSNLT	1062		
QY	1740	IMPPSPRPTFGITWRYRLQTVSLAGVSL,MLLWASLRMDMAKVP,PGGSGTRTETS	1792			
Db	1063	IMPPARPRPTLQCMKQOTLMARSLAHVALQ,LIKIMSSIKFNFD-PDTHPRRXYI	1119			
QY	1800	ET--EITTEILIKRDVGPY,IRFECIRKIICPI-GVDEPREKPTTPORKG-----	1850			
Db	1120	DTPSHDERRRRIIRKEMPPYGOXYEREMET,IELPLYDEDEEDESLSNRGSSSEFSGR	1179			
QY	1851	SSALRPRPEP-----	PKQ-----	GVYIETVABEDELME-	188	
Db	1180	SSSARKRPQOHEFLSLKFGNYPRKSKMAFRSLDNRRATA,IREWWDGYTLKVFBIKDYW	1235			
QY	1884	--IRAF,ERVEKEKAQV-----	BOQAKR,LEOQKPTVIAITSTPSTSTSTISPAO	1934		
Db	1240	KWIAAELEKTA,KKRLELT,TKRAQAKADEBERR,IQOQOQOSVAR,IEY-PMHS---	1294			
QY	1935	KVMVAPISGSV---TTGTMLVLT,TVGSPATVFOON-----	KNFHQFATW,KGQ	198		
Db	1295	RNNV-PYLSGOOORBRP,GNERNGLFLEKYNNSSSVSPAHGYASTPPPGYHQPPNII,IRAG	135			
QY	1984	SNSGVVOQKVLGII,PSSTGTSQOFTFSFO,PTATVITRENTSGSGGTSNSOYITPQ	204			
Db	1354	YNO-----	LPK,ETTS,PFNFOS-RP-VAT,IPPTPOLRAAGADG--VVRAM	136		
QY	2044	IRP,-----	MTVIRP,LOQSTLGA,IIIRTVVMQ,PGAPQVMOIIRGO-----	208		
Db	1397	MTPEKNS,TVNTNS,TPYQO-----	ALNRQOYQ,LOROQOQ,PAVRL,INGHFMGD,ITRGGGR	145		

```

OY 2088 -----PVSAVSAFNTVSTFGGXSLSAT-----STSNIOSS 21260
Db 1452 NPVQMIQRLPQNAALQRFPGSESTEMRYVTAALIPDNDSGEOPPVIPRDPISNFPAQ 151111
OY 2121 ASOPRPQOGVKITMAQLTOLTOGHGNGOGLTVIYIQGGOTTGOLQ-----IPGQTVL 2176
Db 1512 RAQOOHQPOSRPVYVSTPAQMIKRTTOPGVGNH--NVILMKASDGTOMMYLKPGQFGPGTVI- 1568
OY 2177 PGPGQOLMOAMPNGTVQRFLEPLPLATFTATVSTSTTTVSTTAAGTGBGROSKL----- 2230
Db 1569 -STQGRVYVPTKQPLAVQQRDLIT-----ATPQTRKYVRIIPNANGAPRQODHQWKRYYQ 1621
OY 2231 --SPOMOHODK--TLPPAQ-----SSVGPAAKAPQTAOPSARPOPOQOS 2274
Db 1622 ASGPRAMEYMDQSTPRPGGOVRYVLQGGNSGTPNVPNKVSSRGGPRGGLTMOWMOQOQ 1681
OY 2275 PAQEVQTOPEVQF-----QTIVSSHVPSQAO 2301
Db 1682 QHNPAHYDMEDDATGFVNSTTTTQGVYDEQO 1711

```

RESULT 7					
ID	O95ZU8	PRELIMINARY:	PRT:	1713 AA.	
AC	O95ZU8;				
Dt	01-DEC-2001 (TREMBLrel. 19, Created)				
Dt	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
Dt	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	F2EH11.2b protein.				
GN	F2EH11.2B				
OS	Caeornhabdilis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
CC	Rhabditidae; Peloiderinae; Caenorhabdilis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Burlew K.;				
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99069613; PubMed=9851916;				
RA	none;				
RT	"Genome sequence of the nematode C.elegans: A platform for investigating biology."				
RL	Science 282:2012-2018(1998).				
DR	EMBL; Z81515; CAC42289.1; -.				
DR	InterPro; IPR004022; DDT_dom.				
DR	InterPro; IPR001965; znf_PHD.				
DR	Pfam; PF02791; DDT_1.				
DR	Pfam; PF00628; PHD; 1.				
SQ	SEQUENCE 1713 AA; 196824 MR; 178E470017D9AD5A CRC64;				
Query Match 5.7%; Score 848.5; DB 5; Length 1713;					
Best Local Similarity 17.4%; Pred. No. 5.4e-26;					
Matches 434; Conservative 295; Mismatches 619; Indels 1147; Gaps 74.					
QY	4	EEEEEE-----GDGAETDDSEDE-----EDEMEEDDDSDIYEEMEDD---DDASCTE	52		
		: :: :: :: :: : :: : :: :			
Db	169	KROEDIITYMEDSEDSEEBESSDSDEMLEDVOVEEBEEELNLTDIKIEKGIDEENKCY--	226		
QY	53	SSFSRSHSTYSSTPGRRKRPRVRHRSPLNEE---KDLPLEFPKSSEDLMWPNHMINTAI	110		
Db	227	-----PWLDDEPASLKPLEPESODIPPIIASINDAVEI	261		
QY	111	YEVLINFGTVLRLSFREFDEFCALVSQCOTLMAEMHVILAKAVLRBEDTSNTTGPAD	170		
		::: ::: :: : :: : :: : :: : : : : : : : : : : : :			
Db	262	YEILRSYHRTLRIPTPEPFDEFCALISHNNSCIMEAEVMALLRNCLKGSDEEQTHSYTE	321		
QY	171	LKDVSNSTLIYIDGTMWEVYLRYVCSESKEYNHNLYPYOEA-----EDPIGYEVNKI	222		
		: :: :: :			
Db	322	TNNSVNMIHHMDLLYVAELLROYTEA-----YPFADASYRDIAINDVPNVFYVDANI	374		

QY 223 K-----VLQFLVDOFLTTNTIARELMSEGVYQVYDHCRCVCHK-LGDLIC 265
Db 375 QDESFEPFNKHFRLVYLFLMSYRFLLYSSEFKLVNNVNGFQDNENCRVCKSGSGRVA 434
QY 266 CETOSAVYHLBCVPRLEBEDEMOCEVAAHK-VPGVTCOVARIQCNKPIRIHEPIGY 324
Db 435 CTQCEAAHFVEC--SHLPFPE-VLVCNICKKNSAVRGVLPPEDEVADPEP-LRQPIGR 489
QY 325 DSRKKYWFNLNRLIIEEDTENENEKKIMYVSTKVQALIDLCDLDKWEALCKILEEM 384
Db 490 DRYGTYVFIYRLVLOSLDETE---LYYSTVQVLQVLQKDRYTYEKOLCTIRLR 545
QY 365 REEIRHMDITEDLTNRKRGSKSFLAANEILESIRAKK---GDIDNVSPREETDK 441
Db 546 IDEFLQOMALIVEMTSERR-----EALLETWVKROLIGYDFAEATPPQ----- 588
QY 442 NETEDSDAKENREEFDOSEKSDOKTPD---DPEOGSEPTVEGDKGNSVANL 498
Db 589 -----IYLRDSMKRMASILRDCAKQGVKOEVL----- 618
QY 499 GDNMTNATSEETSPESGRSPVCSLSETPDSSNMAEKVASELPQDVPEPNKTCESNNTS 558
Db 619 -----EE--PYEGOSPVKCVQFVEDSI-----LPE----- 641
QY 559 ATTTISIQNLNENSSSELNSSQESAKAADDPENGERBSHTPVSIQEIYGDFTSEKST 618
Db 642 -----SMIGIFPAKLINTFWSGA-----TOEELVQEFV---DI 672
QY 619 GELSPGAGGASGSTHIIIRLRNPDLSKLSQOYAAAHANKLFKECKEVLVANS 678
Db 673 SDNFAPSA----- 681
QY 679 QGELSRLSTKKEVINKNINNYFKLGQEGK---YRVYHNOYSTNSFALNKHQREDHDKR 735
Db 682 -----NLMRMGDEGNQDTFMTYNYYSRNMSESEFLRKKAAQDK 721
QY 736 RHLAKFCLTPAGEFKW---NGSVHGSKVLTISTLRILTLOLENNIDSSFFHPWASHR 791
Db 722 KYMAKSF--AQIDNEDWVAVAKNROFYGDASLCKEIMMTLOQVINIIPDLHRRKPEFA 779
QY 792 ANWIAVOMCSKPREFALALILECAVAPVYMLPIRREFIGTRLHRMTSIEREKEKYK 851
Db 780 KGFDELVAVADDYKLVCLKLDCAVKRTJFMPQWNGSLGOTRLERTIVDQREPNMKQ 839
QY 852 KKEKK-----QEEETMOQATWVYKTFPVYKHQVAKOGEEYRVTG---YQGSWISKT 901
Db 840 QRLKIDADALTKDLDDSFVRVNMKPKWPNY-ILRQGEYTRNAGSGMGMVAVAK 898
QY 902 HYRFVPKLPNGTVNVYKRSLEGTNNMDENMDESDDRKCSRSPKIKIEPDSEKDEVKG 961
Db 899 YVEKMI-----QVPESPKLPPLAVTVEIKTESVSNRK--- 930
QY 962 SDAAGADONEMDISKITEKKDQDVKELLDSDDPKCEEPMEVDDMKTESHVACQESS 1021
Db 931 -----ARLELIVSKIT----- 942
QY 1022 QVDVYVNSEGHILRTSYKKTTSKLDGLBERIKOFTLEEKORLEKIKLEGIGIGKT 1081
Db 943 -----KKRQSG-----GKS 952
QY 1082 STNSKNLSESPVITAKKEGQOSDMROQSPNANNQPEDLIQCSOSDSSVLRMSDFS 1141
Db 953 SKKPFELT-----NGCYSFSCRS----- 971
QY 1142 HTNKLKYKRDVLDVYSIRSEPTKCPKONSTIENDIEKYVSLASRGQEPYTKTKGNDF 1201
Db 972 -----NPNRKC-----YSP----- 980
QY 1202 IDDSKLASADIGTLICKNKKRPLQIESDITVSSKSAHSHSVKSTNDRCATPLSRAMD 1261
Db 981 -----MCNN----- 984
QY 1262 FEGKLGDSSENSTLSSSDTVSIQDSSEEDMIYONSMESISQEFRTREQDVELEPIKC 1321

Db 985 ----- 984
QY 1322 ELVSGESTGNCEDRLPVKTEANGKPKPSQOKLEERPYNKCSQJIKLKNITDKKNENRE 1381
Db 985 -----GYLSAKQAHDEKLEB----- 1001
QY 1382 SEKKQRTSTQINGKDNKPKIYKGECLKEISESRVYSGVNEPVNMINKIIPENDIKS 1441
Db 1002 ----- 1001
QY 1442 LTVKSAIRPFIINGVIMEDFERNSSETKSHLSSDAEGYRDSLETLPSTKSDSTQ 1501
Db 1002 ----- 1001
QY 1502 TTTPASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLSDNFIDENGLPINKNEN 1561
Db 1002 -----SG 1003
QY 1562 YNGESKRKTVITEVTMTSTVATESKTVIKYKGDQVTVSSTENCASVTITTTTVTK 1621
Db 1004 VLGEKK----- 1009
QY 1622 LSTPSTGGSVDIISVKEQSKIVVTTVDSLTTTGGTLVTSMTVSKEVSTRDKVLMKFS 1681
Db 1010 ----- 1009
QY 1682 RPKTRSGTALPSYKRFYTKSTKSLFVLPNDLKLARKGIREV--PYFVYNAKPALD 1739
Db 1010 -----AMPPEIOTFTSKRGKSIYVLQKKILKOMIMGCCQOYVMPGFAGIKSNLL 1062
QY 1740 IMPYSPRPTFGITWRYRLQTVKSLAGVSLMLRLMASLRMDMAKVPYGGGSTRTETS 1799
Db 1063 IMPYAPRPTLDCWKQWQJLNARSLHVAVALQKIIMSSIKFNEFD--PDTHPNRRVYI 1119
QY 1800 ET--EITTEILIKRDVGPYIGIRFEXCIRKIICPI--GVPEYKKEPPTQORKG-----LR 1850
Db 1120 DTPSHDERRRRIIRHKEMPYGQYEREMEIETIPLYDEPEEDESMLNNRKGSSSEFSHR 1179
QY 1851 SALSRLKREPET-----PKOT-----GVIIETVVAEBELTME----- 1883
Db 1180 SSARKKRPQRHEFLSLKFGNTPKKSNAFRSLDNRRAIRREWDVTLKVPFELQWY 1239
QY 1884 --IRAFEREREKAQAV-----EQAKKRLQOKPTVIASTSTSPSTSTSTISPAQ 1934
Db 1240 KIRAEAKTARKLEATRTKAKAKADEERRRIQOQOQSVARIPV--PMHS-----LIPSE 1294
QY 1935 KYVAPAPISGV---TTGTKMYLITKVGSPATVTRQOQ-----KNFQTFATVYKQOQ 1983
Db 1295 RNNV--PYLSQOQORRPNRNGERGFLEKYNSSSVSQAHGYASTPPGYHQDPNIIROAG 1353
QY 1984 SNSGVVOYQOKVLGIIPSSGTSGQTFSPFORPTAVTIRPRTSSGGTSSOYITGQ 2043
Db 1353 YNQ-----LPKRTTSPFNFS--RP-VATIPPTPLRAAGAGD---VYRAV 1396
QY 2044 IRPG---MTVIRPLOSSTLKAIIRPVVAVQPAPOQVMTQIIRGO----- 2087
Db 1397 MTPGNKSTVNTNSTPYPQ-----ALNRQOYQLOROQOQPAVARLNLNGYHFMGTMRGGR 1451
QY 2088 -----PVSTAVSAPNTVSTPGOKSLTSAT-----STSNIOSS 2120
Db 1452 NPSVOMHQRLPQNRALALQPFGESTTEMRRVTEAALIPNDGDECPVJIPRYDPTSNFDAQ 1511
QY 2121 ASQPRPQOQGVKLMMAQLOTLQGHGQOGGLTVIQGGQGTGLOL-----IPQGVYVL 2176
Db 1512 RAQOQHPQSRPYSTRQAMIRTTQPGVKN--NVLKMASDQTMVLKPGQFPPTGYI- 1568
QY 2177 PPGQOLMAAMPNGTVQRFLETPLATATTAJTTTSTVSTAAGTGEORSKL----- 2230
Db 1569 -STQGRVYVYRQPTAVQOROLYT-----ATPGTIVVIRPNANGAPRQODHQVBRVQ 1621
QY 2231 --SPQ-MOVHODKTLRP-----AQSSSVGPAKAQPOTAQPSAPPO-----POTQ 2271

DB 1622 ASGPRAMEYNDQGTPEPGQVRYVLOGNSGTPNVNPPKVSRRGSGRLGIMOMVQOOO 1681
QY 2272 POSPAQ-----PEVQTPREVOQTIVSSHVPSQAQ 2301
DB 1682 QHNPFOAHYMDPDDATGFAVSTTT---EQVPDEQQ 1713
RESULT 8
0960Y3
ID 0960Y3 PRELIMINARY; PRT; 1022 AA.
AC 0960Y3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE ID30146P.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY051776; AAK93200.1;
DR Flybase; FBgn0035118; E(bx).
SQ SEQUENCE 1022 AA; 111667 MW; 1174939B61962E63 CRC64;
Query Match 4.6%; Score 692.5; DB 5; Length 1022;
Best Local Similarity 25.3%; Pred. No. 5e-20;
Matches 294; Conservative 157; Mismatches 440; Indels 269; Gaps 47;
QY 1520 DMEIETSEVKKYSPTSEESNL-----SNPFD---ENGLPINK 1558
DB 6 NVEIETEDSIYTLGNSGNAEDVDMTPGRRKRNOKSKSYIGTKIVLDQTLDKDIPINK 65
QY 1559 NENVNESKRRKTYTEVTMTSTVATESKTYIKVEKDKQTVVSTENCAKSTVTTT 1618
DB 66 -----QNRREPIT-----ARPKREC-----VKYERETENGNERV----- 97
QY 1619 VTKLSTPSTGSAV---DIISVKEQSKYVTTTVDSTLTGTLTSMVSKYSTROK 1674
DB 98 ---YSTSSPRGRVYLLNDAKLYEQA-----VKTEDEK 126
QY 1675 VKLMKFSRKRTSGTALPSYKFKVSTKSKIFVLPNDLKLARKGIREVPYFNVA 1734
DB 127 STITK--KTSYSR---YLLISNFTLHKKKRLSLVLPRELLKALLGKSSNGHHNA 180
QY 1735 KPALDIWYPPSPPTFGITWRIRLQTVKSLAGVSLMLRLILMASLKMDDMAAKVPEGGS 1794
DB 181 KNN-TTWOQCSPLPRTQWSYTSNATSLALQILIMWSLIRMDMIAM--PSTDEK 238
QY 1795 RRETSTETITTEIIRKRDVGPYGFCEYCIKRIICPIVCEPKTPTPORKGSSSL 1854
DB 239 HOVTITTEIVTLELKLRIHSGRKEGTSTLRKKVILPLEPKVREV-TSISGLR---- 293
QY 1855 RKRREPPTQGTGVILETWAAEELEMEIRAFAREREKEKAQAVQO-AKKRLEQOKPT 1913
DB 294 KKRRAASPTEPQITEEMVDEKLEMEIKFGEKQEKARLSAVTRVSRASQLE----- 348
QY 1914 VVATSTSTSTSTSTISPAQKVMAP-IGSVYTGTKNVLTKVGSPTVTFQOKNFH 1972
DB 349 --ASGNGSNTSTNGALGAAGRVQLAPKLEEDVK--EKMEQDLKT--GRAVHQQRK--- 398
QY 1973 QTFATVVKQGSNSGVVQOQVLS-----IIPSTGTS---QOTFSFQ---PRATVY 2021
DB 399 -----LVATGETLRSVTPVKGOVIGSRRAVIVKMPDGTTRIIQOAVTVQVSRTGAGNTAAA 453

QY 2022 TRPNTSGSGTTSNS-----OVIIGPO-----IRPGMTVIRTP-----LOOSTL 2060
DB 454 ASPTVGGSTSTOSNBSPTSPHKVQIIRGPDGKVSXRGILNPGQOLVMPGKRLHVLT 513
QY 2061 GKAIIRTPVWVGAPQVMTQIIRGQPVSTAVSAPNTVSTPGOKSLTSATSTNSIOSS 2120
DB 514 SNSAGGNKMKVPIKPASTSS-----SPAISSAQTTNPTVPIKQIAVHVKNKSNATQSI 569
QY 2121 ASQPPRPQOGVYKLMALQTLTQGHGNOGLTVYIOGGGOTTGQ-----LQLIPOG 2172
DB 570 AS-----SSRYALPLAQI-----KNKLLAQOQOQSTSSPATSSPVOKIYK 613
QY 2173 VYVLPQGOQLMOAMPNTVORFLETPLATATTASTTTVSTTAA--GTGEOROSK 2229
DB 614 VYNTSTSGOTLQOVYQSS-----KLIVGQANQAGKVIISTMAAQOQSTSPVQOQ 665
QY 2230 LSPOMOVHODKTLPPAOSSS-----VGPBK--AOPQTAQPSARPQOTOPSPAQPEVO 2281
DB 666 LVQSPFIQOS-----PQIISMTOQOQIYVGGQRIILSPGQITVYQRNVPGQALOMVQOQIQ 721
QY 2282 TQP-----EVQQTIVSSHVPSAOPHTHAQSSKPYAAASQPSQSNVQSGSPRYVQSP 2333
DB 722 TQOQOQOHHVQOPQOQEVVQSNQIVQSSPSAQTKLVKQLVWQOQSOQTIEERTQIT 781
QY 2334 SQT-----RIRSTPQSLSPGOQOQVOTTSOPPIPIQHTSLQIPSOQPOSQPOVQSS 2388
DB 782 NETGTVQVLPVPSSTLAQQLAGKQLVATVNGQOVIYK-----LGNNQADLVNHI 831
QY 2389 QTLSSQGLTNQVSSSPSPQOLQIQ-QOPQVYAVPQLOQOVVQ--LSQIQSOVVAQIQAO 2446
DB 832 KIQGDSNNAHIVTNSATVAVQANPQTSVPKQOALPQSPQOVVQVQOQIHOQSPNTESG 891
QY 2447 QSGVPOQIKL-----QPLQIQSSAVQOTHOQN-----VTVQASAVQ----- 2485
DB 892 VPTITQOQVLPVTAQVQAPQ-QQALSVESLQNOQPPGYIKCVTAQVLOTETGPRIVLOG 950
QY 2486 -----BOLQVQV-QLRDQOQKKKQOQIEINVTPTSKLIRKELIQOVYMKHNAVIE 2536
DB 951 LVGNDFTAQQLQVQVQVQOQLMKAQESNGKLGIVGPTIYLVQO-----PENAV-- 1000
QY 2537 HLKQKSKMTPAEREENQMT 2556
DB 1001 -QOQPELTPVHQSAHQOV 1019
RESULT 9
Q9W0T1
ID Q9W0T1 PRELIMINARY; PRT; 997 AA.
AC Q9W0T1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG7022 protein.
GN E(BX) OR CG7022 OR CG17135.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kohira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskehn D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003467; AAF47362.1; -
DR FlyBase: FBgn0035118; E(bx).
SQ SEQUENCE 997 AA: 109295 MW: 8106E06CF6A8C7 CRC64;

Query March 4.4%; Score 661.5; DB 5; Length 997;
Best Local Similarity 25.4%; Pred. No. 8.4e-19;
Matches 286; Conservative 151; Mismatches 409; Indels 279; Gaps 48;

QY 1520 DMEITSEVKKVTSPITSEESNL-----SNDFTD---ENGLPINK 1558
D 6 NVEIESTEDSLVTGLNSGNAEDVDMTPEMRRKRNOKSKSVYIGTFVDLDQTLDKRPIPLNK 65
QY 1559 NENVNGESKRTVTTEVTMTSTVATESKTVIKVEKGDQKTVSSTENCAKSVTTT 1618
D 66 -----ONRRFPIT-----ARPKREC-----YKVEREFENGENR----- 97
QY 1619 VTKLSTPSTGGSV-----DIIVKESKTVVTTTVDLSLTGGTLVTSMTYSKESTDX 1674
D 98 ---YSTSSPKRGRVLLNDAKLYRDA-----VKTEDK 126
QY 1675 VKLKKFSRPRKTRGTALPSYRKFTKSTKSIFFVLPNDLKLARKGIREVPYFNA 1734
D 127 STIRK--KPSYSR-----YPLISNFLT HKKRSLLVLPREFELLKLARKGKSTNFFHAA 180
QY 1735 KPALDIPYSPRPTGTWYRIQTVKSLAGVSLMLRLNASLRWMDMAKVPFGSGST 1794
D 181 KNN-TIMQYQCSRLPFTCMYSRTSNATSLSLALQTLRIILMSCLFMDMIK--PSTGK 238
QY 1795 RTESETFETITTEIKRRDVGPGYRGFEYCIRKIICPTGVETPEPTPKRGKLRSSAL 1854
D 239 HOVTDTDFIVLELKLHRSRGTEKTSYLRKRVYIPLMKRYAREV-TSIRSGR----- 293
QY 1855 RPKRPETPKGTAVIETWVAEELELWIRAFERVEKEKAQVEAO--AKRRLDQCKPT 1913
D 294 KRRAESQHPQEPQITEWVEDKLELWEIKFEGEKOKARLSAVTSVAHQLE----- 348
QY 1914 VIAISTSPSTSTSTSPAKWVAP--ISGSVTTGKMTVLTKVGSFATVTPQONKPFH 1972
D 349 --ASGSNSNSTGALGVAGVOLAPKLISDVK--EKMEQOLKL---QRAVHQRK--- 398
QY 1973 QTFAFWKOGSGNGSVQVOOKVLG---ITPSSNGTS---QDFTSQ---PRTAVT 2021
D 399 -----LVATGETITSVTPVKGVIGSRVYIKNPDGTTIRIIQAVTQVSRTGANTAA 453

QY 2022 IRPNTSGSGGTTSSNS-----QVITGPQ-----IRPGMTVIRTPLOOSTLCKAIR 2066
D 454 ASPTVGSGSTSTQSNBSTSTPHKVIIRBPDGVKSVRGINPGQOLVQMP-----DGKLVHL 508
QY 2067 TPVWVQPGAPQ-----QVMTQI-----IRGPVST-----ANSANTYSS 2101
D 509 TTTTSSNAGOGGLFKLKSNTITWVYHFLGKMKKVPKIPASTSSPALSSAQTTN 568
QY 2102 --TP-----GKSLTSATSTNIQSSASOPRPPQGOYKLVFMQTLQTCGHHGUGLTV 2154
D 569 PTPYPIKOLAKVHYKNTKSNATQSIASS-----RALPLAQI-----KNKL 609
QY 2155 VIQGGQGTGQ-----LQIPGVTVLPBPGQOLQMAAMPNGTVORFLFTPLATTAT 2206
D 610 LAQOQOQSTSSSPATSSPVQKIVSKVNTSTSGTLLQOVFVQSGS-----KLTVGQ 661
QY 2207 TASTTTTIVSTTA---GTGEOROSKLSFQKOVHD-KTLPPAOSSSVGPAP--AQOTA 2260
D 662 NAQGGKVTITSTMAQOQSTSPVQOQOLVQSQPIQSPQOISMTOIIVGCGRIILSPQT 721
QY 2261 QPSARPOPOTOPQSPAQEPVOTP-----EVOTTTVSHVSEAOPTHAQSSKPOV 2312
D 722 IYTCNNVQSQALQWVQOQIQOQOQOQHVVYQOQQVYVQSNQIVQSSPSAQTLVQL 781
QY 2313 AAQSQPQSNVQGSFVRYQSPSQT-----RIRSPQSLSPEQSQOYVTTTSPPIQPH 2367
D 782 VYQOQSQGTIEKQITTTDSNETGTQVLVPSNLQAQGLQVATVNGQYIVMP- 840
QY 2368 TSLQIPSOQPOQPOVQOSSSTQTLSSGOTLQVSVSFSRPLQIQ-QOPQVAVPQLQ 2426
D 841 -----LGNQQAQIYAHKIQDGNHNIYTSNATVAVPQANQTSVPKQALPQSP 891
QY 2427 QOVQV-LSQIQSQVYVQAQQAQSQGVPOQIKL-----QLPQIQQSSAVQTHQIN----- 2475
D 892 QOVVYQOQOIHQOQSTNEBSCVTPIQOPVLQAQVAPAQ-QQALSVESLQONPQGTV 950
QY 2476 --VTVQAASVQ-----EQLQRYQ-QLRQDQCK 2500.
D 951 IKCVTAQVQLQFHEGPRIVLQGLVGNDFTAQQLQVLQTVQVQKQLMK 995

RESULT 10
Q9VC00
ID Q9VC00 PRELIMINARY; PRT: 2768 AA.
AC Q9VC00:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG13648 protein.
GN CG13648.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephygroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chumpe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

	Durbin K.O., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagum R.
OY	Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McMahon D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Mosherfi A., Mount S.M., Moy M., Murphy B., Murphy L., Mutzy D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Sidon-Klamos I., Simpson M., Skupski M.P., Smith T., Splier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svyrskas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X., Wang Z.-Y., Wassaran D.A., Weinstock G.W., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yen R.-F., Zaveri J.S., Zhao M., Zhang G., Zhou Q., Zheng L., Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.). RA RT "The genome sequence of Drosophila melanogaster." RL Science 287:2185-2195(2000). DR EMBL; AE003750: AA656376.. - DR FlyBase; FBgn0039257; CG13648. DR InterPro; IPRO01007; vWF_C. DR Pfam; PF00093; vwc_1. DR SMART; SMO0214; VWC; 4. SR PROSITE; PS01208; VMFC; 1. SQ SEQUENCE 2768 AA; 294032 MP; CA929A1774E4684 CRC64;
Query Match	4.2%; Score 621.5; DB 5; Length 2768; Best Local Similarity 18.9%; Pred.No.1.le-16; Matches 584; Conservative 448; Mismatches 1197; Indels 865; Gaps 133;
Oy	78 PLLEKDP--PLFPKSSEDLMPPNEIMNVAIYEVRNFGVILRISPRFFDFCAA 134 ::: : : : : : : : : : : Db 84 PIITCEPVAVDVPYHSPEPTGLSIPERK--GCSIEEKFEYGAGVPSNPKNPCELCYC 140 ::: : : : : : : : : : : Oy 135 IVSOECCTMAEMHVLLKVALREEDTSNTTGPADLKDSYSTVFIDGMTPVLRYV 194 ::: :: : : : : : : : : : : Db 141 INNOIKRCWQ-----ECTIH-YDG-----CLPLY 163 ::: : : : : : : : : : : Oy 195 CESDEKYHHVLPYOAEADYPVGVENKIKVLOFLVDGFITTN----- 236 ::: : : : : : : : : : : Db 164 NKGS----CCPVRYSCHD---ENELD---FMDSFTTTVTTVRPYTGTILA STMTP 209 ::: : : : : : : : : : : Oy 237 -----IARELMSEGV-IQYDDHCRCVKHL-GDLIC---CE-----TCGANVHL 275 ::: : : : : : : : : : : Db 210 PTSTDICHGEIFADGAASLRGNACBHCXMGDIIVCAVOEECVPMANANGSKRAMPA 269 ::: : : : : : : : : : : Oy 276 ECVCRRPLEVEPRDEMOCCEVCVAHKRVGYTDV-----AEIOCKNRPYIRHEPIGRRSR 329 ::: : : : : : : : : : : Db 270 BEGCCPSNATVCDBSDSTTEIYLTTPESASTSPAKGHAALRKEDVDLDGHIDDKKKE 329 ::: : : : : : : : : : : Oy 330 KYMFLNRIILIIEDPENENEKKIWYSYTKVOLAEILDICDKDYW---EAELCKILEMKRE 386 ::: : : : : : : : : : : Db 330 TATIPSaelgSGEI DEEEKND--kATTVAAYQTd-EKFSPESPESSTAIGISR-DKR 383 ::: : : : : : : : : : : Oy 387 EIHRMDITEDILTNRKARGSNKSFIAANEIIESTIRACK---GDIDNYKSPETEKOON 442 ::: : : : : : : : : : : Db 384 -----IDPSSSTEESKESSTAAAEEDIKVIYTPTEPEGSGEDVPKPQSQEIKEIT 434 ::: : : : : : : : : : : Oy 443 ETET--NDSKAENKREE-----PEDQ-----SLKDSDKTRPDODDPQG 479 ::: : : : : : : : : : : Db 435 EDLIKIVSTSAPAKSAPEEVVTAATSAPTDEDVKTPTTAGTISEEBEGKRP--ABEG 491 ::: : : : : : : : : : : Oy 480 KSSE-----PREVGKGNSVANIGDNNTINASETESPERS-----P 518 ::: : : : : : : : : : : Db 482 SSEBKDVKVTAPALETDEAKPTSAPV---ASIDEKGGPKRSSBSGBDELDTKPTTAPT 548 ::: : : : : : : : : : : Oy 519 VGCISETIPPSSNMARKEYVASELLPODVP-EPNKTCSSNTSATTTSIOPNLENSNSSSEL 577 ::: : : : : : : : : : :

Dd	549	AGATSASESE0DEGK - STEAPL SVDDLEPAKPTESSEASG-----EGEDVAKET	599
Qy	578	MSOSEASAKADDPEDENERESHTPVSI - DEEYI - GDPTKEKSTGELSESPGAKGASGT	635
Dd	600	TPAGASIASAGEELVYKTTTPAGERSSGDELYKGTTPABESSSEDE-----	648
Qy	636	RITRLRNPDSKLSOLKSQVAAAHAANKLEKEG--VLVNSOGELISLSTKKEVIM	693
Dd	649	-LTFVTPPAGEPSVAGEEELAKETTPAGEPSIAGEEELVYKTTPAGE--SLIAGEEELVYK	705
Qy	694	KGININNYKLGQBEKRYVYHNOYSTNSFALNKHQHEDHDKRHLAKFCLTPAGEFKN	753
Dd	706	-----VTPPAGESSEGEELIK-----VTPPAGESSE 734	
Qy	754	GSVHSGKVL-----ISTRLTITTOLENNIPSSFFPHNASHRAMWIKR - VOMCSKRE	806
Dd	735	GDEELVYKSTPAGERISEGEDVYKATSPAKSDIGVAKEPLETAEVPAEVEDPAK--	792
Qy	807	FALALALIECAVYKVVMLPTWREFLGHTRLHRRTSIEREKEXYKKEKQOEETMOQA	866
Dd	793	---TTPIAEEEPPIAGTPIPTGIS-----GEEELIKGTTPOLEBOPRISBS	839
Qy	867	TPWXTPEPVKQVWKQGEERVATGVGQWMSIKTHVYFVFKLPONTNVN - YRKSLEGT	925
Dd	840	TEV-----PVAED-----DLSSITSASIASSTGCV	865
Qy	926	KNNMDENDESDKRRCSRPKIKIIBDSKED-----EYKSDAKGADON	971
Dd	866	QDAASETTSAPARAGDKDEAATVTPAQOKDEVEGQATDLPEDVYQSTTAKTTTE	925
Qy	972	EMDISKITEKKDQVYKELBDSDKPCKEPEMVDDMK-----	1010
Dd	926	QPKERSSTEADAEILEVYTSPPAKQEVPAEBADHDHDEEDVQATDLPKSDIGPV	985
Qy	1011	--TESHVANCOSQOVVANNVSEGFHL - RTGYKKTKTKSKLDGLERIKOFL-----E	1061
Dd	986	VDTEATTQOPETSDETATDKPPSYVYLPVYSEVPSSTAYVNDNFETEKPLPPSGEQ	1045
Qy	1062	EKORAEIKLEGGIGK-----IGKSTNSKMLSESVYITRAKEGOSDSNROOSFNA	1115
Dd	1046	SEPLPAMDLPAGIPGEGDCLVEBKTYANNTYPAAPADCVSK--CISLIAVACQMECK	1103
Qy	1116	NNDOPE-----DLQCSOS--DSSVLRMSPSHITKMLPKDRLVDVDSIRSEPIKC	1166
Dd	1104	LPENEKCTVAAADLLDGCCPTIODESTESAEDBESTAK--PDNKIDEDVSEISTE -EI	1160
Qy	1167	PKONSIENDIEEK-----VSLASGQEPSTSKTNGDFLDSOKLASADI	1213
Dd	1161	PRDVIMPTGITEQPLSHVKPDEEIQPVTYPAQFDSTIAKADKP - IDES-----	1210
Qy	1214	GLLICKNNKPLIOESPTIVSSKSAVHSSVPRKSTDRATPLSRAMFEGKLGCDSESN	1273
Dd	1211	-----AEDKKPIGSEED-----SKPIDESEEDKKVE-----	1238
Qy	1274	STLENSDPTVSIQDSSSEDMIV-----ONSNESTISEQFRTREOD-----VEVLEPLKE	1322
Dd	1239	---ESAEDKCPVEDESEKEREPLTPVIPASIEKESKPEDEKTEADFAAPTEOPATPPA	1295
Qy	1323	LYSGSTENCCDRLPVGTENGK--KPSOOKKLEERVYKNCSDOI--KIKNTDKN	1376
Dd	1296	QIADIAEKEVDOKLATTSAVPSGEDELKPADEKKRE-----TQOIPDAELIPASTOPE	1349
Qy	1377	NENR---ESEKQCORST--FOJNGKDNKPKIYLKGECLKEISRRVYSGVKEVKNVI	1430
Dd	1350	STTELPYVDLKKKEEDSTGTEAPESDKVEEPTASITENIEESDKFTYIAPPKISAS	1409
Qy	1431	NKIIP--ENDIKSLTVE-----SAIRPFIINGDY-----MEDERNSESTKSHL	1474
Dd	1410	DETEPTAEDLVATPTFERISEEFVSTKRAVGGPLPLTAPAGEKKPVAAETSTADT	1469
Qy	1475	ISSDDEGNYDLSLETLPTKRESDIQTTP - SASCPESNSVNOVEDMEIETSEYKATTS	1533


```

Db 1470 STEPSAEVEKASGETSESDNEIDAGASSTPVPVSADEKTPSTKTEVTE---ADDKFTTV 1526
QY 1534 SPIT--SEESN---LSNDFIDENGLPI-----NKNENVNGESK-----RKT 1570
Db 1527 APLAGEEESNLPKPLQDIFEEBA-PVAATTAAPSKDD---GQOKVEVEBEKIEDGQRP 1582
QY 1571 VTEVTMTSTVATESKTVIKVEGDKQTVVSTENCASVTVTVTYTKLSTPGGS 1630
Db 1583 IEDETSTPTS-----SENEIEPESDRATTIAPSKEE--PSEPSTGAPTDEPAEPSTADP 1635
QY 1631 VDIISVKEBSKVYTTVYDLSLTGTLVTSMTVSKVSTRKVKLMKFSRKKTRSGT 1690
Db 1636 ESDSE--KETPESEVPTTAPA---GEKIPTSSITPDEEPTATSAVARDEDEVEKETSTP 1690
QY 1691 ALPSYRKFTYKSTKKSIFVLPNDLKLARKGIREVPYFNNAKALDIWPPSRPPTF 1750
Db 1691 EIPTOAPASSEEDENS-----STDQIPS-----EVP-----EKKP----- 1720
QY 1751 GTTWRYRIQTVNSLAVSLSMLRLMLASLMDMAKVPYPPG---GSTRETSETETITTE 1807
Db 1721 -----ETPAQTPEBGDIVGATVAPTTSDVEYPPVQ 1749
QY 1808 IIKRDNVGYGIREYCIRKIIICPIGVPETPKTPPTPQPK--GLRSALKRKRPTKQTC 1866
Db 1750 RLPE-----EVLAEIIPQSTETGIRKODETTAAFSIDRK-- 1783
QY 1867 PVIETWVAEELEMEIRAFARVERK--EKAQVBO-----QAKRLRLOOKPTVIATST 1919
Db 1784 -----EYVTEIDEATVAPISKEDEKPTPEEKPVQOKPGEPESEBEKEKIEDOVST 1839
QY 1920 TSPTSSTTSTISPAOKVMVAPISGSVTTGTKMVLTTKVGSPATVTPQONKFNHOTATWV 1979
Db 1840 EGVSTSEASEAGSTES-----SEEVKPTSEGVAKKPED----- 1873
QY 1980 KQOGSNS-GVVOVOQVNLGIIISSTG---TSQOTFSFQPRKATVTVIR--PNTSGSGCTTS 2034
Db 1874 KQPSSTAQAAPVETIPISTELPQADQDKPTISEAPVSDDETSKAPSEDEKIPSVSG----- 1927
QY 2035 NSQVINGPOIRPQMTVIRTPAQSTGKALIRTPVWOPGAPQOV--MTQIIRQOPSTAV 2093
Db 1928 --EVEGPEV-----TTASQAABEDELKTPASEPSBSTKVETEVOKREDETKAD 1977
QY 2094 SAPNVTSSTPGOKSLTSATST-----SNIOSSAQPPRPQOGVKLTMALQTLQTOGHG 2148
Db 1978 EMPESVTQVSDVATSTAPAGAGDIEKDEQATYASP--EBEEIKRPTIAPAAEIPQ----- 2031
QY 2149 NQGLTVIIGQGGTQGLIIPGVVLP--GFGQOLMO--AAMPNGTVQRFLETPLATTA 2205
Db 2032 -----PSEKEPVDE--QEVESGTATAPAESDGPIDETIAPATSGPIDE-----ASTA 2076
QY 2206 TTASTTTTSTTAAGTGROSKLSPOKOVHODK-----TLPPAOSSSVGCPAKAOPOT 2259
Db 2077 APTKESTIVASGA-----SP--AVHDEIKVDVTTTQPVADKEKEVAPRODETKT 2123
QY 2260 A-----QPSAR--POPOTOPQSPAPOPEVOQTPEVQT--QTTVSSHVPs---EAOPTH-- 2304
Db 2124 SIDVSTDSPTAODEKQDTEAPVAPTIVSSPTADSAASTPTVAVPSVPELDTGPMMD 2183
QY 2305 --AQSCKPOVA-AQSPQSNVQGSQPVRY--QSPSOTRIRPS-----TPSOLSPGOQSO 2353
Db 2184 IMSQRTAAPTADCAASTFEDQDAFVTVSPDAEKTVPVSPAPQSDKTPSSSAPADDAE 2243
QY 2354 VQTTTSQ-----PIPIQHTSLQIPSGOPQSOPOVOSTQTL-----SSGOTLNQVS 2401
Db 2244 IRATATPLDDNKIRATVAAQOTDGDVATAPALDEBKIQTTAALDEEKIPSTAAPLDDK 2303
QY 2402 VSSPSRQOLQOPQOVIAVQLOQOVYLSQIQSQVAAQIQAOQSGVPOQIKLOLPIC 2461
Db 2304 IPAPVSPVVFVDEVPSESEKPAVSEYDE-----ESTEPVHVDVETSTDEPTSDAKLRPTS 2358
QY 2462 IOQSSAVQTHQIQNVYTVAA-----SVQOQLOVQOOLDQOQKQKQOQIETIVNTPSKL 2516
Db 2359 APATPESPATEAETVETTAPELEKEVPEKATEQPELEKETPEKATEQPELEKETPEKA 2418

```

```

QY 2517 LIKVEI-----IQQVYMKHNAVIEHLKQKSMTPAER-----EENORMIV 2557
Db 2419 TEOPELEKETPEKATEQPELEKEVETKATQEPQESVDEKTPPEVYVAPSLDSTDEDE----- 2474
QY 2558 CNOVMKYIILDKIDKEKQAAKKRRREESVQKRSKONATKLSAL----- 2601
Db 2475 -----ESVSEESASAKKDKNKETEBEDTDKHEEPVAVVSEIQPQSEAAVPTTG 2525
QY 2602 --LEFKHEQRLAEILKRALDLQDQIEVOEELKRLKIKKEDLQQLQATVAAACPP 2659
Db 2526 HPLFPH---LASSTTTTPAVDDR-----VGE-----DENTTVKLSSTTTSTTESP 2570
QY 2660 VT-----PVLPAAPPAPPPSPPGVOHTGLSTPTLPVSAQKRRREKSSSKSKKKM 2715
Db 2571 VTSAPSTTVYASQOQOPPIPTPPYG--HA-----PEYDEIDEVEVROPGTCRYAGKL 2620
QY 2716 ISTTSKETKKDKTKLYCICTPYDESKFYIGCDRCQNMWYHRCVGIQS-----EABL 2767
Db 2621 YVSAQOQIPRDDPCDFCFC-----FRSDIICLOQSCPPPIAGCHEEP 2661
QY 2768 IDEYVCPQO-----STEDAMTVLTP 2788
Db 2662 ISGFCCPPEYECVPVMAAVLNTTSTTTTSTLTP 2695

RESULT 11
Q9NMJ17
AC Q9NMJ17 PRELIMINARY; PRT; 5476 AA.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Split ends.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11)
RP SEQUENCE FROM N.A.
RA MEDLINE=20157049; PubMed=10655223;
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,
RA Suh C., Voas M., Williams A., Rubin G.M.;
RT *A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
RT Protein Kinase Signaling Pathway That Interact With the yan Gene of
RT Drosophila Identifies Split ends, a New RNA Recognition Motif-
RT Containing Protein."
RL Genetics 154:695-712(2000).
DR EMBL: AF184612; AAF26299.1; -.
DR HSSP: P09651; 1HA1.
DR FlyBase: FBgn0016977; spen.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 3.
DR SMART: SM00360; RRM; 3.
DR PROSITE: PS50102; RRM; 3.
SQ
SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match 4.1%; Score 613.5; DB 5; Length 5476;
Best Local Similarity 19.5%; Pred. No. 5,1e-16;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
QY 305 DCAAEIQKNNPYIRHHEPIGDRSRKRYWPLNR-----RLIIEDOT 344
Db 1855 DTVADKAEKKNQ--RHBEKKEKREKEDLRKQVEBEKDRKAQOEEREKDKAAEER 1911
QY 345 ENENEKKIWWYSTKYVOLAEILDLCDKYWEAEILCKILEEMREIEIHHMDITE-DLTNKKAR 403
Db 1912 ERREREKKAQEDRK-----KEREREKLR--KEDRQKQEKREKREKDLREKQ 1958
QY 404 GSNKSFLLAANEELIESIRAKKGDIDNVKSPETEKQNETENDSKDAEKNREEPDQSL 463

```


Db 1959 -----RERDNREKELRDOKLREKEMREKQREKELAREKQDOREREHR 2000
 QY 464 EKSDOKTPDDPE--OGKSEEPTE-----VGOKGNSVANLGONTNATSEETSPS 513
 Db 2001 EKESRRANDYEOBGRGRRELSSYOKSKMDIAGEASSLTAIDCOKNKANADTLAOGT 2060
 QY 514 EGRSPVCLSETPDSSNMAEKVASELPODVPEPNK---TCESSNTSA-----TTTST 564
 Db 2061 PGASP-----STP--SDNTPKERSRKLRSNRPALHKKRLLSQOESNHSAGGSGSGSSH 2113
 QY 565 QPULENSSELNSSQSEAKAADPENGEBES--HTPVSIOEIVGDTSEKSGELS 622
 Db 2114 QIHHEVDYKRIREMSONISVHSSNORLNRDRSKERHSSSFED-----KNSSSHIS 2166
 QY 623 ESPAGGASGSPRIIRLNRNPSKLSQLKSGOYAAAHANKLFKCGEVLVYNSGEL 682
 Db 2167 RPHCGGSSSSSKHNHR-----RDKHHQGSASSETNSSI-----EYVVDPIISQTK 2214
 QY 683 SRLSTKEVIMKGNINNYFLQOGKYRVYHNOYSTNSFALNK-----HOHR----- 729
 Db 2215 HNLTSEELQSHQ-----PKREKERHPSSHANSSSRHKSRRDHHHREKRRHSA 2267
 QY 730 -----EDHOKRRRLAKFCILT---PAGEFKNGSVHSGKVLITSLRLITIOLENNIP 779
 Db 2268 ESTNTDEHPTPOQHNPRIIRISAAGSGAGELSSAATWTSQKLLHGHRRSVERKSSRGS 2327
 QY 780 SSFPHPMASHRAWITAKAYOMCKPREPALALALECAVPVVALPIWREFLGHTIRLHRM 839
 Db 2328 DEGHSSSKSLRAKLMLLSASDDTDDA-----SKHNST 2362
 QY 840 TSIERE-----EKERYKKKKOEETMOATWKYTFPVKHQVMKQGEERYVTGYG 894
 Db 2363 FDIPODCPNVMYDKVARCKNMROAEKKIKAKPS-----QLOKSORAKKRSTSYDG 2417
 QY 895 WS---WLSKTHVYFVYKLGNTVWNRKSLSEGTKNMNMENMDSOKRKSRSRKKIKIE 951
 Db 2418 DSDTEFEDROH-----RNSGSSSFHGR--YPLGSSSDDDDEETHORRIS----- 2460
 QY 952 PDSEKDEKVSADAKGADONE-----MDISKITEKKDOQVKELLDSODKPCKEEPMYVD 1006
 Db 2461 -----SBSDAHGGODNOGASTLADANRVROMO--QNLRLICDGDGS---SED--EIR 2506
 QY 1007 DDMKTESHVNCQESSQYDVVNVNSEG-----PHLRTSY-----KKTKSSKLDGLL 1051
 Db 2507 RNMYKHHFEGKRNNSSTRIASDSESQOPARDLTIKQEHPIAQAOLIKRQOLSEDEQFK 2566
 QY 1052 ERRIKQITLEKOKLEKIKILEGKIGKIGKSTNSKSKLSESPVTTKAKKEGQOSMSMOEQ 1111
 Db 2567 SRHDSNSISIER---KLTKERET---KTELGDFTNSSEYTYGKLEKXSPETRRKHKK 2618
 QY 1112 SPNANNQPEDLIOGCSQSDSS-----VLRMSDPS---HTTNKLTYPKRVLDVSI-- 1159
 Db 2619 SKRR-----LKSSSTADTSAOQPLVMTPLRPSIFDVHSSSECKTKDNPNFDLTEC 2670
 QY 1160 -----RSPETKCPKONSTENDIEEYKVDLASRQOEPTKTKYGNDFIDSK 1206
 Db 2671 SSIPLEISAGERRRKHKEKKEKREKLRNMTAEATVPN-----SEPTMDT-----SSE 2716
 QY 1207 LASADDIIGTLCKNKKPLIOESPTIVSSKSAHSSVPMSTNDRDPTLSR----- 1288
 Db 2717 KLSKEEHRRLKSKSSKSMONSCNTKIYNSGA-H---PSTSPSLPATPTPSAPSTAOTSK 2772
 QY 1259 -----AMDFEGLGDSSENSTLENSSDTVSIODSEEDMIVONSENESITSEOFTRREDOVE 1314
 Db 2773 RGEKMEFIRGITIDEESESQPEPBAETNKDIPBS-----VSTTGPIVSALAQTYYKQ--- 2824
 QY 1315 VLEPLKCELVGSESTGCEDELPAVGTBEANGK---KPSQOKKLEERPVKNCDOIKLKN 1370
 Db 2825 --EP-----STPNS-----KNEEAHTOLTIVHEPEOQOOLER---SRLSGSSSSS 2864
 QY 1371 TTDKKNENRESEK---KQORTSTFOINGDNKPKIYLKGECLKELISEKSVYVSGNNEPK 1426
 Db 2865 HADREHRRERREKREKRSQREOQNOIHOKSSK-----VETK 2902

QY 1427 VNNINKIIPENDIKSLTVKESAIRPEINGDVIMEDENERNSSETKSHLLSSDAEGNYRD 1486
 Db 2903 VDDDNISVDMEAGRALEAQ-----LMSDRTKRISB----- 2933
 QY 1487 SLETLDPSTK---ESDSTOTTTPASCPESNSVNOV-----EMELETSYKVKVTSPTT 1537
 Db 2934 --EATPTATATVRSMDTDFERFSDN--EDNNSVMTKQGVKSEQOEQHKSKDKKKRRSK 2990
 QY 1538 SEESNLSNDFIDENGLPIKKNENVNGESKRKTIVTEVTTMTSTVATESKTIVKVEGDK 1597
 Db 2991 EERQEKLL-----QOORSLRNVASTSSAPPTPKKLYVNOVAASK 3031
 QY 1598 QT-----VVSSTENCANKSTV-----TTTTVTKLSTPSTGSDIISVYES 1640
 Db 3032 HADLOLDAKHISSPPYCKPSPSLPCLIGDDDDALHTPKAKPTTPSSRGNDGLTPREKR 3091
 QY 1641 KTV---VTTTVTDSL------TTGTVLVS----- 1662
 Db 3092 RLSPITPKPTLIANSSTLSTQSAETPVSGCTVSSSALATTPSTAGVSAAPGLDNP 3151
 QY 1663 -----MTVSKEYST-----RDKVKLKFGRPK 1685
 Db 3152 TSASAOCKKKESTIPGFDGOLDRISBASAVOSISAEFNSILLDNINADEKPIFVASPPRA 3211
 QY 1686 TRSGTALPSYRKFEVTKSTK----- 1707
 Db 3212 TKPLDKLEESKSVHTIISOETESAVSALLGESFGTSTTDYSLDGMEMSVALETPTL 3271
 QY 1708 FVLPNDLKLKLAK-----GIREVPYFNNAKPAIDWYPPSP-----RPFPGIT 1753
 Db 3272 VIAEPDEEALAKAIETAGEPASILEEP---EMEBERAEBDPPEAEIESEPEVEVL 3327
 QY 1754 WRRL--OTVKSILAGVSLMLRLMASLRWMDMAKVPGGSGSTRETESETTEITELIKRR 1812
 Db 3328 DPBELKAVOSLKHEDM-----DIKADTPOSEDLIDIT--DIFENDE---AD 3372
 QY 1813 DVGPYGIREFEYCIKIKIICPIGVETPREPTPOKGLRSSALRPKRPETPKOTGPVILFT 1872
 Db 3373 SSGP-----SLKIDETVQSSSSPEKSISNNS-----PPREFRANIDI-- 3409
 QY 1873 WVAEELELMEIRAFARVYKEKAQAVEQAKKRLBOQR-----TVLATSTTSPTSS 1926
 Db 3410 -----PNVESQPKLSNESTPOQSVITKPLFDLPKTVPRVGLPPSPVKIE 3453
 QY 1927 TSTISPAQKVVAPRISGV---TTGFKVLTTRKVGSPATVTFQOKNFEHOFATVWVQ 1982
 Db 3454 PPTISKIQPLQPVQVULPAPHSSTGSIANSVINLDSNVSSCSNTSAASATLASASA 3513
 QY 1983 QNSGVVOYQOKVILGILPSSGTISQOTFTSFOP--RTATVTIRPNTSGSGGTTSN---SOY 2038
 Db 3514 SISFGSPITASQNM---POASTPKQGPITPOQAIRFOSILMOPTTISIPEDQTHFAVPOM 3570
 QY 2039 ITGPQI---RPG--MIVITPRLQOSL---GKAIIPT-----PVMVOPGAPQ 2078
 Db 3571 VLSPOHNPQOPGTVMVIGIAPSPHSLPSRPGKAVOSRLVGOQLSPVGRMVPQ--SPQ 3629
 QY 2079 VMTQIIRGQPVSTAVAPNTVSTPGQSLTASITSNIOSSA---SOPRPOOGVQKLT 2135
 Db 3630 ---QVQOTOQHALLITSPOSNSNISPLASPTTYRLVSSNSPTTKVNSYQORNO----- 3680
 QY 2136 MAOLITOLTOGHGNGOGLTVVIOGOGOTTGLOLIPOGVTVLPRPGOOLMOAMPNGTVOR 2195
 Db 3681 -----VPOQSPKSVAVAVQTTPOLMTI---PLQKMPRIQVPHNPT-- 3717
 QY 2196 FLPTPLATATTAATTTTSTTAAGTGEOROKSLSPQOMVHODKTLPRQOSSSVGRARA 2255
 Db 3718 --ISKVTVTVOPOQATQOVASPPPLGS-----LPHKKNVH---LNAHON-----QO 3759
 QY 2256 OQOT--AOPRARPOPOQOPAPAEVQOTQOTVSSHVPESEAOPTIAOSKPOVA 2314
 Db 3760 QPOVIAMKTAHONQHMQOTMHOQMTORQOHMO-----QOQLHGOSO--QITS 3805


```

QY 895 WS---WISKHYRYFVPLRGNTNVYRKSLSTGKTKNNMDENDESDKRCSPKKIKLE 951
D 2475 DSDTEFEDKHQ-----RMSGSSFFHGR--YPELSSDDDDDETHQRIKIS----- 2517
QY 952 PDSEKDEVKSDAKAGADONE-----MDSIKTEKKDQVKELLDSDSRCKEEMFYD 1006
D 2518 -----SSDDEHGHGQDNOGASTLADANRVQRQW-QNLRRLCQDDDS--SED--EIR 2563
QY 1007 DDMKTESHVANCQESQOVVYVNNSEG-----FHLRTSY-----KKTKSSKIDGLL 1051
D 2564 RNVMKHSHFEGRNSNSTRIASDESQOPAPDLTIKOEHPHAPAEIKREBOLSDDEOKER 2623
QY 1052 ERKIKQFTLEKQRLKIKLEGIKIGKTSNKSNSPEVITAKEGCQSDSMROEO 1111
D 2624 SRHDSNSSLEER---KLTREBEL---KTELDFYNSEFYTTGGLKKEYSPTRKKHKK 2675
QY 1112 SPANANDQPEDLIQCSQSDSS-----VLKMSDPS---HTTNKLYPKDRVLDDVSI-- 1159
D 2676 SKRR-----LKSSSTADTSAAGTPPLVMTPLTPSIFDVHSSSECKTKFDNFDLKTGEC 2727
QY 1160 -----RSPETKCPKONSIENDIEEVYSDLASGQEPYTSKTKGNDFIDDK 1206
D 2728 SSJPLEISAGERKKHKKREKKREKLNMTATVPN-----SPTTNDT-----SSE 2773
QY 1207 LASADDIGTLICKNKKPLIOEESDTIYSSSKSALHSSVPKSTMDRATPLSR----- 1258
D 2774 KLSKEEHRKLRKSKSKSMNSCNTKIYNSSGA-H---PSTSPSLATPTIASASTAOTSK 2829
QY 1259 ----AMDESKLQCSSESNTLENSDPTVSIQDSSEDMIVONSISISQFTREDOVE 1314
D 2830 RGEDEKMEFIFGIISDEEESQFPEQAEFNKDIIIPSS-----VSTGTPIVSAALQTYKO-- 2881
QY 1315 VLEPLKCELVSGSGTCEDRLPYKGTAEANG---KPSQKKLEERPVCXCDQJIKLN 1370
D 2882 --EP-----STPNS-----KNEFAHIQLTVEHPEQOOLER---SRLSGSSSSS 2921
QY 1371 TTDKNNENRESEK---KQORTSTQINGDKPKIYLKGECLKEISERVVSGNVEPK 1426
D 2922 HADREHRRRKRREKREKREKQREQOQIHQSKK-----VETK 2959
QY 1427 VNNINKIIPENDIKSLTVKESAIRPFIINGDIVIMEDFERNKSEFKSHLLSSDAEGVYRD 1486
D 2960 VDDNSVDMDEAGRALAEQ-----LMSDFDKPISE----- 2990
QY 1487 SLEPLPSTK---ESDSTQTTTPSPASCPESNSVNOV---EDMELETSEVKKVTSSPIT 1537
D 2991 --EATPPTATYRSDMTDVRFSDN-EDNNSVDMTKQGVKSEQEQHKKSKDKKKKKRKK 3047
QY 1538 SEESNSLNDFTIDENGLPIKNNENNGESKRKTVITEVTMTSTVATESKTVIKVEKGD 1597
D 3048 EEKQEKLL-----QOORRESLPMNVASTSSAPPTPGKILTVNVQVQASK 3088
QY 1598 QT-----VYSSFENCASKTV-----TTTTTTVKLSPTSGSGVDIISVEQOS 1640
D 3089 HADIQLDAAKHISPPVCKPSPSLPCLIGDDDDALHTPKAKPTTPSSRGMDGLTPSKREK 3148
QY 1641 KTV-----VTTVTVDLSLT-----TTGGTLVTS----- 1662
D 3149 RLISPIPTKPTLIANSSTLSTQSAETPVSSGTVISSALATPTTSSAAGVSAAPGLDNP 3208
QY 1663 -----MTVSKEST-----KDKYKIMKFSRPKK 1685
D 3209 TSASAKKKKESFTPGDQLDRISSESAVQISIAEFNSTSLDNIADDEKIPVASSPRA 3268
QY 1686 TRSGTALPSYKREYTKSTKK-----SI 1707
D 3269 TKPLDKLEESKSRVTSIOEETESAVSALGSGFTSTTDYSLDGMDESSVNFLEPTL 3328
QY 1708 FVLNNDLKLARK-----GGIREVPYFNNAKPALDIWYPS--RPTFGIT 1753
D 3329 VIAPDEEALAAKAIEIAGEPASIIEEP---EMEPREAREPDPDEAEIESEPVVEVL 3384
QY 1754 WRYRL-QTVKSLAGVSLMLRLIMASLRMDMAKAVPPOGGSTRIETSETELTTTEIILKR 1812

```

```

D 3385 DEFEINKKAVQSLKHBDM-----DIKADTPQSERDQIDPT-DTEENPDE---AD 3429
QY 1813 DVGPRGIREYVCIKRIKICIGVPEPKETPTPQKGLASSALRRPRRPOTGVIET 1872
D 3430 SSGP-----SLKIDETVOSSSPEKSTSNNS-----PTEPANIDI-- 3466
QY 1873 WAEERBELMEIRAEVEREKERKAQAVEQAKRYLEOKP-----TYIATSTSPST 1926
D 3467 -----PNVESQPKLSNESTPQPSVITKLPFLDPRKTVPAGLPSPVKIE 3510
QY 1927 TSTISPAQKVMVAPISGV---TTGKMWLTTKVGSFAYVTFOONKMFHOTFAIWMQ 1982
D 3511 PPTSKLOQPLVOQVOTVLPARHSTGSGISANSVINLDSNVISSCSNTSASATASASA 3570
QY 1983 QNSNGVVOQKVLGIPSTSTSOQTFTSPQR-PTAYVTIPMNSGGSGGTTSN---SQV 2038
D 3571 SISFGSPITASQNM---PQASPPKQGPRTPOQALRQSLIMOPRTISIPQOTPHAVQ 3627
QY 2039 ITGPOI---RPG---MTVIRTPLOOSTL---GKAIIRT-----PVWVGARQ 2078
D 3628 VLSPOHNPQPGTYWVGIRAPSRHSLHSPGRVAQSLVGLSPVGRMVSQD-SFQ 3686
QY 2079 VMTQITRGVSTAVSAPMTVSSTPQOKSLGATSTNIGSA---SQPRPQCGVKLT 2135
D 3687 ---QVOQTOOQHALLTSPQSSNISPLASPTTVLSSNSPRTSKVNSYQPRMQ----- 3737
QY 2136 MQLQTLQGHGNGGLVTVIOGOGOTTQOLIPQGVVLPGRQOOLQOAMRQTVQR 2195
D 3738 -----VPOQPSKVAEYOTTQOLMI---PQOKMPTGVPHNPT-- 3774
QY 2196 FLETPPLATTATSTTTTTVSTTAAGTGEORQSKLSPQWQODKTLPRPAOSSSVGRAPA 2255
D 3775 -IISKVVTQVPOQATQSQVASSPRLS-----LPHKNVH-----LNAHQN-----Q 3816
QY 2256 QROT-AQPSARQROQVQPSRQAPREVOQOTVTVSSHNPSAOPRNAOSSKROYAA 2314
D 3817 QROVAKMTHAQHQAQOQRMQOFRMQOTQROQNMQ-----QOQLHGSO---QITS 3862
QY 2315 OSQPOQNSQVQSPVAVQVQSPQTRIRPSTPSQLSRQ-QSOVQTTTSQPIQPIHPSLOIP 2373
D 3863 APOQHQAQOHAQOQOQHQAQOQOHNQOHLNQOHLAQOHPRTQKHAQOQOFRMQ-QIOQHOSQ-- 3918
QY 2374 SQGQPOQPOVQS---STQTLSSGQTLNOVSVSPBRQLOIQOQPR-QVIAVPLQOQ- 2428
D 3919 -QHVOQOQOHAQOQOHLQSOQHOSQOQLOMQ---QHQAQOQOQLOQLOKLOMQHGRPOOQOKS 3973
QY 2429 -----QVLSQLOSOVVAQIOAQOQSGVPOQIKQLEPIQIOSSAVQTHQIONV 2476
D 3974 PQGVGHGSGTSLFASQOHNQOLPAR-----QVPOQ---QHPQOLSHSSPCKPMTLVSV 4024
QY 2477 -VTVQASVQEOLORVQOQLRDOQKKQO-----QIEINV--NTP 2513
D 4025 NQGVQPRALITRVGSHSQPQOQOQLPHQOSSGHPHQOLSSPGANLPIQTLPLMYONTP 4084
QY 2514 SKLLIKVEIT-QQVQ---VAKHNAVYENHLKOKKSMPPAREKQNMIVCQWIKYLTDKI 2569
D 4085 -KIIIVQOHNVAQOQVPPROTQGNALHYPONQGDSTPRGH----- 4123
QY 2570 DKEEKOAKKRRKRESEVEQKRSKONATKLSAL-----LFKHEQOLRAELIKKRALLDK 2622
D 4124 -VEPTPAMSAQKISESVSVIRPTPTTGLAVISANTVGSILTEENLIKISQPKRODELRI 4182
QY 2623 DLQIEVO-----ELKRDIKIK-----EKDLMLQATAVAAACPPVTYVLPARPR 2670
D 4183 DSK-EVDSQVSAKEVAVINDSVIKKLDTPLASKAKAKAVEQOAL-----AP 4226
QY 2671 PPSPPPPPGVQNGHGLSTPTLPAVASOKRKRKEEKSSSKKKKM 2715
D 4227 APIPNQPGNQ--SMAQETALPTTSMVNSNHDHDETETRQL 4269

```

RESULT 13

09NH1
ID 09NH1 PRELIMINARY; PRT: 5554 AA.
AC 09NH1;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE SPLIT ends long Isoform.
SPEN OR CG18497.
GN
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Embryoides; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RA Kuang B., Wu S., Shu Y.-A., Luo L., Kolodziej P.;
RT "Split ends encodes large nuclear proteins that regulate neuronal cell
fate and axon extension in the Drosophila embryo."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF221715; AAF34661.1; -
DR HSSP; P09651; 1HA1.
DR FLYBase; FBgn0016977; spen.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.
SQ SEQUENCE 5554 AA; 59918 MW; 4037E27833D0C622 CRC64;

Query Match 4.1%; Score 613.5; DB:5; Length 5554;
Best Local Similarity 19.5%; Pred. No. 5.2e-16;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;

QY 305 DCAVEIQKPKYIHEPIGYRSRRKYWFLNR-----RLIEEDT 344
DB 1906 DTAQKAENQ---RHEKKKKRQERREKDLKQYREKKDKRAQOEERKKKEEK 1962
QY 345 ENEWEKTKWYSTVQALIDLDKDYWEALCKILEMREBEIHRMDTE-DLTNKR 403
DB 1963 EREKREKQOEERK-----KEREERELRE--KEQRDKQEKKEIREKDLREKQ 2009
QY 404 GSNSFLAANBEILESTRAKKGDIDNVKSPETFEKDKNFTENDSKDKNREFEDSL 463
DB 2010 -----REKDNREKELDKDLKREKREKQERKELHREDDQERHR 2051
QY 464 EKSDDKTPDDPE--QKSEEPTE-----YDKGNSVANLGDVTNATSEETSPS 513
DB 2052 EKESRRAMDVEQSGRGMRMELSYQSKMDIAGEASLAIIDQNHKRNAMDTIAGT 2111
QY 514 EGRSPVGLSTPDSSNAEKVKASELPQDVPPEPNK--TCESSTSA-----TTTSI 564
DB 2112 PGASP-----STP--SDTTPKERSRKLSRNSPVLRLHRRRLSSQSEHNSAGGGCGSSH 2164
QY 565 QPNLENSSSSELSSQSESAKADDPENGERS--HTPVISOETIVDTSEKSTGELS 622
DB 2165 QIHHEDEVKRIEMNSQNIIVSHSSQRLNDRDSKEHKSSEFKED--KSSSHIS 2217
QY 623 ESPAGKAGSGSTRILTLRLPDSKLSQLKSOYAAAHNAKMLEKEGEVLYVNSQEI 682
DB 2218 RPHCGGSGASSSKHNHR-----RDKHKGSGASSIETNSSI---EYVVDISOTK 2265
QY 683 SRLSTKKEVINKGINNYFKLGQGGYRVYHNOYSTNSPALNK-----HQR- 729
DB 2266 HNLMTSEELQSHQ-----PKREKREHFSHANSSSSRHKSRRDHNHREKKRHSVA 2318
QY 730 -----EDHKRRHLAKFCLT-----PAGEFKWNGSVGSLVLTSLRLITTOENNIP 779
DB 2319 ESTNTDEHTPQOHNPRIISAAGSGSAGELSSAATNTSSGLHQHRRRSVERKSSGS 2378
QY 780 SSFHPNVAHRAWIKAVQCSKPREFALALALECAVKVYVMLPIRFLGHTRLHRM 839
DB 2379 DEGHSSSKSLRAKLMLMSADSDTDIDA-----SKKSI 2413
QY 840 TSIERE-----EKERVKKKEKQEEETMOATWVKTYPVKHQQVKKQGEERYVTGYG 894

DB 2414 FDIPDDCPNVSMTDKVKARSCNMQRAEKKIKAKFS-----OLKQSRAKKRSTSYDG 2468
QY 895 MS-----WISKTHYRVVPLPGNTVNYRKRSLSEGTNNNDENNDESDKRCSPAKIKIE 951
DB 2469 DSDTFEDRQH-----RNSGSSSFHR--YPLGSSDDDDDEETHORIS----- 2511
QY 952 PDSEKDEVKSDPAAGADONE-----MDISKITEKKDQVVELDSDSDKPKCEPMEDV 1006
DB 2512 -----SDSAEIHGGDNGGASGLAANRYRQW-QNLRRLCGGDS--SED--EIR 2557
QY 1007 DDMKTESHVNCQESSQVDVNVNSG-----FHURTY-----KKTKSKSLDGLL 1051
DB 2558 RVMKSHFGRKNSNSTRIASDESQSPAPDLTIKQHPILAPAOEIRQOLSDSEQRFK 2617
QY 1052 ERRIQFLLEQRLEKRIKLEGIGIKGTSTNSKNLSSEPVITKARQCGSDSMRQO 1111
DB 2618 SRHDSNSSIEER---KKTEREI---KTELGFYNSSEYTYGKLAEYSPEPRKKHK 2669
QY 1112 SPNANDQPEDLIQGCSDSDS-----VLRMSDPS--HTTNKLYPKRYLDVSI-- 1159
DB 2670 SKRR-----LKSSSTADISAAQTPLVMPPLPFSIDVHSSSCCKTKFQDFDLKTEC 2721
QY 1160 -----KSPETCPKQNSIENDIEEKVSDLASRGOEPTKSTKGNDFFIDSK 1206
DB 2722 SSIPLEISAGERKKHKKERKKREKRLRMTEATVFN-----SPTTNDT-----SSE 2767
QY 1207 LASADDITLICKNNKPLIQESDITVSSSKSALHSSVPKSTGNDDATPLSR----- 1258
DB 2768 KLSKEERHLKSKSKSKMDNSCNKTYNSSGA--PSTSPSLPAPTPASPSTAQTSK 2823
QY 1259 ---AMDEGKLGCDSESNSTLENSSDTVSIQDSSSEEDMIQVONSNESTSEOFRTQDVE 1314
DB 2824 RGEDMEFIFGLISDEESQPEQAETNKDILPSS-----VSTTPRIYSALQTKQ--- 2875
QY 1315 VLEPLKELVSGESTGNCEDRLPVKGTGANCK---KPSQKLEERYPNVNCSDQIKLN 1370
DB 2876 --EP-----STPNSS-----KNEBAHIOLTVHEPEOQOOLER---SRLSGSSSSS 2915
QY 1371 TTDDKNNNRESEK---KQORTSTFOINGDKNRKIKYLKGCKLKEISRSRVNSVNPBK 1426
DB 2916 HADREHRRKREKRRKRSKQEOQNIHQSSK-----VETK 2953
QY 1427 VNINIKIIPENDIKSLTVKESAIRPFIINGDVIMEDFENRSEETKSHLLSSDAGNYRD 1486
DB 2954 VDDDSVMDDEGRALENO-----LMSDFDKPISE----- 2984
QY 1487 SLETLPTK---ESDSTQTPPSASCPSNSVNOY-----EDMEIETSEVAKVTSPIT 1537
DB 2985 --EATPSTAAATYRSDMTDVFRPSDN--EDNNSVDMTKQGVKSEQOQHKSCKKKRRSK 3041
QY 1538 SEESNLSNDFIDENGLIINKNENNVNGESKRTVYTTVTMTSYATSEKVIYKYEKDK 1597
DB 3042 EKKQKRL-----QOQRESLPNVASTSSAPPTGKLTIVNOQASK 3082
QY 1598 QT-----VVSSTENCAKSTV-----TTTTVTYTKLSTPSTGSGVDIISVKQOS 1640
DB 3083 HADLOLDAKHHISSPVVCPSLPCILIDDDDDDLHTKAKPTTPSSGNGLTPSRKRP 3142
QY 1641 KTV-----VTITVDTSLT-----TTGGTLVYS----- 1662
DB 3143 RLISPIKPTPIANSSTLSTQSAETPVSSGTVYISSALATPTTSSTAAGVSAAPGLDMS 3202
QY 1663 -----MTVSEYST-----RKVYLAKMFSRKK 1685
DB 3203 TSASAQCKKSESFIPGFDQGLDRISASAVOSISAEFNSTLNDNIADKPIVAVASPPRA 3262
QY 1686 TRSGTALPSYRKVEYTKSTK-----SI 1707
DB 3263 TKPLDKLEESKRVITISOETIESAVALLGESFGISSTTIVSLDGMDSVNELETPTL 3322
QY 1708 FVLPNDDLKLARK-----GGIREVPFNYNNAKPAALDIWPPSP-----RPTFGIT 1753

Db 1912 DTADKAENQ---RHEKKEKREKREKDLRKOVEREKDKRAQOEKREKEDKAKAEK 1968
QY 345 EBNENEKITYSTKYVLAELIDCLDKDYAEIJCYLEEKEEIRHMDITE-DLTNKAR 403
Db 1969 EBEREKKAODEKREK-----KEREERELRE--KEQDKEQKEKREKREKREKEQ 2015
QY 404 GSKSKSLAANELELESTAKKGDIDNVKSPFTEKDKNTENDSDAKENREEFEDSL 463
Db 2016 -----HERDNREKELDKDLREKEMREKREKELHREKQOREHER 2057
QY 464 EKDDOKTDPDDPE--QGKSEEPTE-----VGDGKNSVANLGDNTTNNATSEETSPS 513
Db 2058 EKEQSRADVDYDEQEGKGRNELSTYOKSKMDLAGASSILATIDCQHNKENAMDTAOGT 2117
QY 514 EGRSPVGLSETPDSSNMAEKVASELPODVPEPNK--TCBSSNTSA-----TTS1 564
Db 2118 PGASP-----STP--SDNTPKERSKRLSRNSPYRLHRLRLSSQESNHSAGGGSCGSSH 2170
QY 565 QPNLENSNSSLNSQSOSAKAADPENGERS--HTVYSIOEELYGDTSEKSTGELS 622
Db 2171 QJHHEDYVRKIRIEMNSQNIYHSSNORLDRDSEKSKSSFEKED-----KNSSSHIS 2223
QY 623 ESPGAKGASGSTRITRLRNPDSKLSQLKSOQVAAAHANKLFKEKEVLVNVNGEI 682
Db 2224 RHHGCGGSSASSKHNH-----RDKHNQKGSASSTETNST--EVVVDPISTQK 2271
QY 683 SRLSTKEVIMKGINNYEKLQEGKYRYVHNOYSTNSPALNK-----HOH----- 729
Db 2272 HNLNTSEELQSHQ-----PKREKEREHFSHANSSSRHKSRODHNHRRKHHVA 2324
QY 730 -----EDDKRRHLAKCLT---PAGEFKNGSVHSGKVLTLTLTLTTOLENNIP 779
Db 2325 ESTNTEDEHTPODHNHRRISAAGSGASELSAATNTSGKLHNOHRRSVERRKSSRG 2384
QY 780 SSFHPNMAASHRANWIKAVOMCSKPREPALALALECAVKPVMLIRWREFLGTLHRM 839
Db 2385 DEGHSSSSSLRAKMLMLSSADSDTDDA-----SKKHST 2419
QY 840 TSIERE-----EKEVKYKREKOEETEQAOTVWYTFEPVKHGVKOKEEYRVGYG 894
Db 2420 FDIPODCPVNMVDKYKARSKCMQRAEKKIKAFKS-----QLKQSRKKKRSYDQ 2474
QY 895 WS---WISTHYRYRPLKPGNTNWKSLBSTKNNMDENDESKRCSRPKKIKIE 951
Db 2475 DSDTEFEDHO-----RNGSSSFHGR--YPLGSSDDDDDEETQORRIS----- 2517
QY 952 PSEKXEVGSAKAGADNE-----MDISITEKKDQVYKELDDSDKPCKEPMEVD 1006
Db 2518 -----SDSDHEHGQDQCASTLADANRVQMO-QNLRRLCDGDD--SED--EIR 2563
QY 1007 DDMKTESHVNCOESQVDVNVNSEG-----FHLRTSY-----KKTKSSKLDGL 1051
Db 2564 RVMWKSHSHGKRNSTRILASDESGQAPADLTIKQEHPIAQLKREQLSDDEOKFK 2623
QY 1052 ERRIKOTLEBKORLEKILBEGIKIGTSTNNSKNLSEPVITAKBCCGSDSMRQO 1111
Db 2624 SRHDSNLSIEER---KLKTEREI---KTELGDFFNSSEYTYGKLKEYSPETRRKKHK 2675
QY 1112 SPANANDQREDLIGOSOSDSS-----VLMSDPS-----HTNKILXPDRVLDDVSI-- 1159
Db 2676 SKRR-----LKSSTADTSAOQPLVMTPLTPSIIFDVHSSSEKTKFDFNDLKTETC 2727
QY 1160 -----RSPETKCPKONSIENDIEEKVSLASGQEPSTKRYKGNDFIDSK 1206
Db 2728 SSIPLEISAGERKKHKEKREKREKRLNMTAEVPM-----SPTINDT-----SSE 2773
QY 1207 LASADDIGTLICKNNKPLIOESDVTIVSSSKSLHSSVPKSTINDRATPLSR----- 1258
Db 2774 KLSKEERHRLKRSKSKSMDSNCSNTKIYNSGA-H---PSTSPILPATPTASASTAQTSK 2829
QY 1259 ----AMDFGKLGCDSESNSTLENSDYVSIDQSSSEDMIVONSNSISIQFTRQDVE 1314
Db 2830 RGEDEKMEFLIGLISDEESQFPEOATNKDILPSS-----VSTTGPVSAALQTYKO--- 2881

QY 1315 VLEPLKCELVSGSTGNCEDRLFYKGTENGK-----KPSQKKLEERPVKKSDQIKLN 1370
Db 2882 --EP-----STPNS-----KNEEAHIQLTVEHEPEOQOLER---SRLSGSSSS 2921
QY 1371 TTDKNNENRESEK-----KGORTSTFOINGDKNPKRYLKGELKELSEERVYSGVNEPK 1426
Db 2922 HADREHRRREKREKREKRSQREOQNIHOKSK-----VETK 2959
QY 1427 VNNINKIIPENDIKSLTVKESAIRPEINGDIMEDENERSSTKSHLLSSDAEGNYRD 1486
Db 2960 VDDDNVDNDEARALEAQ-----LMSDDTKRIS----- 2990
QY 1487 SLETLDPSTK---ESDSTQTTTPASCPESNSVNOV-----EDMELETSEVKKVTSSPIT 1537
Db 2991 --EATPSTATYRSDMTDVFREFSDN--EDNNSVDMTKQGVKSEQOQHKKSKDKKKKKRSK 3047
QY 1538 SEESNLSNDFIDENGLPIKNENVNGESKRKYVITEYMTSTVATETSTYKVEGDK 1597
Db 3048 EERQEKIL-----QOQRRESLPRVASTSSAPPKGLTVNVOASK 3088
QY 1598 QT-----VSSSTENCAKSTV-----TTTTTVTKLSTPSGSGVDIISVKEOS 1640
Db 3089 HADLODAKHISSPPCKPSPSLPCLIGDDDDALHTPKAKPTTPPSRGNDGLTPREKR 3148
QY 1641 KTV---VTTTVDLSLT-----TTGGTLVTS----- 1662
Db 3149 RLSPITPKPTIANSSTLSTQSAETPVSGTVISSALATPTTSRAAGVSAAPGLDNP 3208
QY 1663 -----MYVKEKST-----RDKVKLKKFSRPK 1685
Db 3209 TSASAOCKKKEPIPEPDGOLDLRISAVOISIAEFNSTLNDIADDEPKIIVASPPRA 3268
QY 1686 TRSGTALPSYRKFPVTKSTK-----SI 1707
Db 3269 TKPLDLKEBSKSVTTISOETESAVSALGSEFGTSTTYSILDGMDMSVMELEPTL 3328
QY 1708 FVLNDLKLKLRK-----GGIREVPYFNANRPAIDIMYPSP-----RPTFGIT 1753
Db 3329 VIAPEBEALAKAJETAGEPASILEEP---EMEPERAEEDPOPEARIESEPVAYEL 3384
QY 1754 WRRL-QYVSLAGVSLMLRLMASLRMDMAKVPBGGSSTETETSETITTEIIRK 1812
Db 3385 DPBELKAVOISLHEDMM-----DIKADTPQSERDLOIDT--DTEENDE--AD 3429
QY 1813 DVGPRGIRFEYCIKLIICPIGPEETPEPTPOKGLRSSALPKRPREPKOTGPVIEET 1872
Db 3430 SSGP-----SLKIDETVOSSSPKSTISNNS-----PIPREANIDI-- 3466
QY 1873 WVAEELELMEIRAFARVEREKEKAQAVEQAKKRLBQKP-----TVIATSTTSPTSST 1926
Db 3467 -----PNVESQPKLSNESPPOPSVITKLPLFDLPKTVAPGLPPPVKIE 3510
QY 1927 TSTISPAQVMAVAPISGV---TTGTKAVLTKVSGPATVTVQONKNHQATVATWVKQ 1982
Db 3511 PPTISKLOQPLVOPVTVLPAHSTGSGISANSVINLDSNVISCSNMSAASATASASA 3570
QY 1983 QNSNGVVOVQOKVLGLIIPSTGTSQOTFTSFOP--RATVAVIRBNTGSGGTTSN--SOV 2038
Db 3571 SIFSGSPASQNAM---POASTPKQCPITPOQAIRQSLIMOPTTISIPQTHFAVPOH 3627
QY 2039 ITGPQI---RPG--MTVITPLQOSTL---GRAIIRT-----PVAVOPGAPQ 2078
Db 3628 VLSRQSHNPOQFTYVWIGIARPHSLRPGVAGVQSRVLQGLSPRGVMSQD--SPQ 3686
QY 2079 VMTQIIRGQPVSTAIVASAPNTVSTPGOKSLTSATSTSNIOSSA---SOPRPOGQVKLT 2135
Db 3687 ---QVOTQOQAHLLTSPSSNISPLASPTTRVLSSNSPTTKVNSYOPRNO----- 3737
QY 2136 MAQLQLQGHGNOGLTVVIOGOGOTGOLOLIPGVVULPBGQOLMAAMPNGTVQR 2195
Db 3738 -----VPOQSPKSVAVAVOTTPOLMTI---PLOKMTPIQVPHHPT-- 3774

QY	2196	FLFPLPLATPATASTTTTITVTSTAAGTGEOROSKSLSPMOYQDKTLPRAPOSSSVGRPKA	2253
Db	3775	-TISKVTVVPOQAOTOSVASSPRLS-----LPRKNVH-----LVANOH-----QQ	3816
QY	2256	QPCP-AQPSARPRPOTQPSAPREVOGTQEVOTQTTVSHVPSAOPHTAASSKPPVAA	2314
Db	3817	QPVYIAKTHANOHOQNHQOQTMHOQMOIQROHNM-----QQQLHGSQ--QITS	3862
QY	2315	QSPQPSNVGQSPVRYVQSPSOTRIPRSTPSQSLSPG-QSOVOTTTISQPIPIQPHSLQIP	2373
Db	3863	APQOHMOQOHQAQOQOQHNNHQHLMQOLQAQOQPTQKHQAQOQOQPNQ-QIQOHQSOQ--	3918
QY	2374	SGQSQSQSQPOVQVQ---STQTLSSGQTLQNVQSVSPSPRQLOLQIQOPQ-VLAVALQLOQO-	2428
Db	3919	-QHVOYQOQNAQOQHSLSQOQHOSQOOLNQ-----QHQAQOQOQLOQLOKLODMHGPQOQKS	3973
QY	2429	-----VQVLSIQSQSVQAQIQAGSQGVPOQIKLQPLPIQSSAVQTHQIQNV	2476
Db	3974	PQGVGHLAGSTSPISAPQOHNHSLPR-----GVPGQ---QHPOOLSHSPCKPMTLVSV	4024
QY	2477	-VTVQAASVQEQQLQRYOQLLDQOQKKKKQO-----QLEINV--NTP	2513
Db	4025	NQGVQPRPILIRGSHSQSPNQOQOQLRHQOSSSGHPHQOLSSPGANLPQLQPLNYQTPR	4084
QY	2514	SKLLIKVEII-QQKV---VKKHNAVIEHLKKKSMTPAREENQHMIVNQMYKYLIDKI	2569
Db	4085	-KIIVQHQIVANQNVPRPQVQGNAHYHPQNGKQSDRPPH-----	4123
QY	2570	DKEKQQAKKRRKESEVQKRSQKQMTKLSAL-----LFKHKEQLRAELIKKRALLDK	2622
Db	4124	-VEPTPMAQKTSQSESVIATPTPTPTGLAVISANTVGLSILNEEIKISQPKQDELQO	4182
QY	2623	DLOLEVO-----BELRDKRIKK-----EKDLMQATQAVAAPCPVPTVLPAPPAP	2670
Db	4183	DSK-EVSDVMSAEKVNDIVIKLDTPLASKAKRAKAVEMQAI-----AP	4226
QY	2671	PPSPPPPVQVQHTGLSTPLPVASQKRRKREKQSSSKSKKKM	2715
Db	4227	APINPQPGNQ--SMAQETALPTTSMNVNNSNDHDEDETFETROL	4269
RESULT 15			
Q206C3			
AD	Q206C3	PRELIMINARY;	PRT: 5533 AA.
IC	Q206C3		
DT	01-MAY-2000 (TREMBlrel. 13, Created)		
DT	01-MAY-2000 (TREMBlrel. 13, last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, last annotation update)		
DE	SPEN RNP motif protein long isoform.		
GN	SPEN OR CG18497.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
NCBI_TaxId=7227;			
RA	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ISOL;		
RA	Wisllette E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,		
RA	McGinnis W.;		
RT	"Spem encodes an RNP motif protein that interacts with Hox pathways to		
RT	repress the development of head sclerites in the Drosophila trunk."		
RL	Development 0:0-0(2000).		
DR	EMBL; AF188205; AAF13218.1; -.		
DR	HSSP; P09651; IHA1.		
DR	FLYBase; FBgn0016977; spen.		
DR	InterPro; IPR000504; RNA_rec_mot.		
DR	Pfam; PF00076; rrm; 3		
DR	SMART; SM00360; RRM; 3.		
DR	PROSITE; PS0102; RRM; 3.		
SO	SEQUENCE 5533 AA; 597114 MW; AFG60606DEDF269 CRC64;		

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:31 : Search time 23.5095 Seconds

(without alignments)
5128.638 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MVSEEEEDGDAETQDSE.....KLGFKASRSHNNLQSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	4082	27.3	810 1	FALZ_HUMAN
2	476	3.2	3726 1	TRX_DROME
3	439.5	2.9	3256 1	K167_HUMAN
4	439.5	2.9	3924 1	ANKK2_HUMAN
5	407.5	2.7	3828 1	TRX_DROVI
6	406	2.7	2492 1	ATRX_HUMAN
7	406	2.7	4377 1	ANKK3_HUMAN
8	401.5	2.7	1367 1	AMYH_YEAST
9	391	2.6	5179 1	MUC2_HUMAN
10	387.5	2.6	3911 1	AKA9_HUMAN
11	374.5	2.5	2116 1	MYS2_DICDI
12	373	2.5	2476 1	ATRX_MOUSE
13	371	2.5	1589 1	PHP_DROME
14	367	2.5	3969 1	HRX_HUMAN
15	365	2.4	2453 1	NCR1_MOUSE
16	364	2.4	2349 1	TPR_HUMAN
17	362.5	2.4	2843 1	APC_HUMAN
18	358	2.4	2464 1	MAPB_MOUSE
19	356.5	2.4	3866 1	HRX_MOUSE
20	354.5	2.4	2035 1	HRCL_HUMAN
21	353.5	2.4	2035 1	HRCL_HUMAN
22	352.5	2.4	3562 1	PCVY_CHICK
23	352.5	2.4	3210 1	CENF_HUMAN
24	350	2.3	3329 1	BRC2_MOUSE
25	348.5	2.3	2459 1	MAPB_RAT
26	348	2.3	1140 1	YMG6_YEAST
27	341.5	2.3	1875 1	MCP1_YEAST
28	339.5	2.3	2440 1	NCR1_YEAST
29	338	2.3	1790 1	USO1_YEAST
30	337.5	2.3	4385 1	YF73_CAEEL
31	335	2.2	2090 1	HFCL_MESAU
32	334.5	2.2	2468 1	MAPB_HUMAN
33	332.5	2.2	3726 1	ABF1_MOUSE
			1957 1	YD86_SCHPO

34	331	2.2	1531 1	NFT5_HUMAN	O94916 homo sapien
35	331	2.2	3358 1	PCVY_MOUSE	O62059 mus musculu
36	329	2.2	2869 1	RBP1_PLAIVB	O00798 plasmodium
37	329	2.2	3130 1	DPO2_HUMAN	O60673 homo sapien
38	328	2.2	1679 1	Y109_YEAST	P40457 saccharomyc
39	328	2.2	1863 1	BRC1_HUMAN	P38398 homo sapien
40	326.5	2.2	3703 1	ABF1_HUMAN	O15911 homo sapien
41	326	2.2	2230 1	GOC4_HUMAN	O13439 homo sapien
42	325	2.2	2663 1	CENE_HUMAN	O02224 homo sapien
43	325	2.2	2845 1	APC_MOUSE	O61315 mus musculu
44	324	2.2	1781 1	AKAC_HUMAN	O02952 homo sapien
45	323.5	2.2	2717 1	ZEP1_HUMAN	P15822 homo sapien

ALIGNMENTS

RESULT 1
FALZ_HUMAN
ID FALZ_HUMAN STANDARD; PRT; 810 AA.
AC Q12830;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fetal alzheimer antigen (Fetal Alz-50-reactive clone 1).
OS FALZ OR FACL.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95347245; PubMed=7621746;
RA Bowser R., Giambrone A., Davies P.;
RT "FACL, a novel gene identified with the monoclonal antibody Alz50,
is developmentally regulated in human brain.";
RL Dev. Neurosci. 17:20-37(1995).
CL
CC -i- SUBCELLULAR LOCATION: DETECTED IN BOTH THE CYTOPLASM AND NUCLEUS
OF CELLS IN THE DEVELOPING CORTEX. IN THE ADULT BRAIN, IT WAS SEEN
ALMOST EXCLUSIVELY IN THE NUCLEI OF NEURONS OF THE NEOCORTEX. IN
THE BRAINS OF ALZHEIMER DISEASE PATIENTS, THE PROTEIN IS LOCALIZED
IN A SUBSET OF AMYLOID-CONTAINING PLAQUES.
CC -i- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN THE FETAL BRAIN.
CC EXPRESSION IS MUCH LOWER IN ADULT BRAIN AND IS HIGHER IN
NEURODEGENERATIVE DISEASES.
CC -i- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL: U05237; AAA97522.1; -.
CC TRANSPAC: T04682; -.
CC Genew: HGNC:3581; FALZ.
CC MIM: 601819;
DR InterPro: IPR004022; DDT_dom.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF02791; DDT; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS50016; ZF_PHD_2; 1.
KW Nuclear protein; zinc-finger.
FT ZN_FING 251 298
FT DOMAIN 4 9
FT DOMAIN 30 35
FT DOMAIN 42 46
FT DOMAIN 657 660
FT DOMAIN POLY-TYPE.
FT POLY-GLU.
FT POLY-ASP.
FT POLY-ASP.
FT POLY-ALA.

SQ SEQUENCE 810 AA: 91799 MW: F7E2C992FE5BB96D CRC64:

Query Match 27.3% Score 4082: DB 1: Length 810:
Best Local Similarity 98.7%: Pred. No. 6.1e-139:
Matches 777: Conservative 4: Mismatches 2: Indels 4: Gaps 1:

```

QY 1 MWSEEEEDGAEETODEDEDEDEDEDDDDSDYPEEMEDDDDDASCTESSFSRSHST 60
   |||||
DB 1 MWSEEEEDGAEETODEDEDEDEDEDDDDSDYPEEMEDDDDDASCTESSFSRSHST 60
   |||||
QY 61 YSTPGRRKPRVHRPRSPLEEKDIPPLEPKSSSEDLAMPNEHINVAIAYEVLNFGTV 120
   |||||
DB 61 YSTPGRRKPRVHRPRSPLEEKDIPPLEPKSSSEDLAMPNEHINVAIAYEVLNFGTV 120
   |||||
QY 121 LRIISPRFDFCAIYASOSQCTLMAMHVVLLKAVIREDTSTFTGPDILKDSVSTLY 180
   |||||
DB 121 LRIISPRFDFCAIYASOSQCTLMAMHVVLLKAVIREDTSTFTGPDILKDSVSTLY 180
   |||||
QY 181 FIDGMTWPEVLRVYCSDEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTTNARE 240
   |||||
DB 181 FIDGMTWPEVLRVYCSDEYHHVLPYQAEADYPYGPVENKIKVLQFLVDQFLTTNARE 240
   |||||
QY 241 ELMSSEVIQYDHCRCVCHLGLDLCCECTSAHYHLECVKPRLEVEPDEMOCEVCAHKV 300
   |||||
DB 241 ELMSSEVIQYDHCRCVCHLGLDLCCECTSAHYHLECVKPRLEVEPDEMOCEVCAHKV 300
   |||||
QY 301 PGVTDCAVEIQRKRPYIRHPEIGYDRSRKRYWFLNRLLIEEDTEENKIKIYVSTKVO 360
   |||||
DB 301 PGVTDCAVEIQRKRPYIRHPEIGYDRSRKRYWFLNRLLIEEDTEENKIKIYVSTKVO 360
   |||||
QY 361 LAELIDCLDKDWEALCKILEEMREIHRHMDITEDLTNKAAGSNKSLAANEILLES 420
   |||||
DB 361 LAELIDCLDKDWEALCKILEEMREIHRHMDITEDLTNKAAGSNKSLAANEILLES 420
   |||||
QY 421 IRAKKDIDNVSPETEEDKNTENDSKDAENKEEFEDQSLERKSDDKTPDDPEQCK 480
   |||||
DB 421 IRAKKDIDNVSPETEEDKNTENDSKDAENKEEFEDQSLERKSDDKTPDDPEQCK 480
   |||||
QY 481 SEEPTEVGDRGNSVANTLDNTNATSEETSPSEGRSPVGLSETPDSSNMAKKVASEL 540
   |||||
DB 481 SEEPTEVGDRGNSVANTLDNTNATSEETSPSEGRSPVGLSETPDSSNMAKKVASEL 540
   |||||
QY 541 PODVPEPKTCGSSNTSATTTISIOPLNENSSSELSNQSSESAKAADDPENGERESH 600
   |||||
DB 541 PODVPEPKTCGSSNTSATTTISIOPLNENSSSELSNQSSESAKAADDPENGERESH 600
   |||||
QY 601 PVSIOEIVGDTSEKSTGSELSPGAGGASGSTRITRLRNPDSKLSQLSQVAAAA 660
   |||||
DB 601 PVSIOEIVGDTSEKSTGSELSPGAGGASGSTRITRLRNPDSKLSQLSQVAAAA 660
   |||||
QY 661 HEANKLFKEGKEVLYVNSGGEISRLSTKKEVIMKGINNYFKLGEGKRYVHNQSTNS 720
   |||||
DB 661 HEANKLFKEGKEVLYVNSGGEISRLSTKKEVIMKGINNYFKLGEGKRYVHNQSTNS 720
   |||||
QY 721 PALNKHQHRHDHDKRRHLAHKFCLPAGEFKWNSVHGSKVLTSTPLRLTTQLENNIS 780
   |||||
DB 721 PALNKHQHRHDHDKRRHLAHKFCLPAGEFKWNSVHGSKVLTSTPLRLTTQLENNIS 780
   |||||
QY 781 SFEPHFW 787
   |||||
DB 781 SFEPHFW 787
   |||||
QY 777 TSLHPSE 783
   |||||
DB 777 TSLHPSE 783
   |||||

```

RESULT 2

TRX_DROME STANDARD: PRT: 3726 AA.
ID TRX_DROME STANDARD: PRT: 3726 AA.
AC P20659: Q27255: Q27327:
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila melanogaster (Fruit fly).

```

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
RT in Drosophila, encodes a protein with zinc-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
RN [2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=7924996;
RA Sedkov Y., Tiliib S., Mizrokhi L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
RT Drosophila embryogenesis than is the Antennapedia complex, correlating
RT with a bithorax-like expression pattern of distinct early trithorax
RT transcripts.";
RL Development 120:1907-1917(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RT between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokhi L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
RT directly regulates the region-specific homeotic gene fork head.";
RL Genes Dev. 8:2478-2490(1994).
CC - FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.
CC - MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
CC ZINC.
CC - SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC - SIMILARITY: CONTAINS 1 SET DOMAIN.
CC - SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31617; AAA29025.1; -
CC EMBL: Z50152; CAA90514.1; -
CC EMBL: Z50152; CAA90513.1; -
CC EMBL: Z31725; CAA83516.1; -
CC EMBL: Z31725; CAA83515.1; -
CC PIR: A35085; A35085.
CC HSSP: P20393; 1A6Y.
CC TRANSFAC: T00850; -
CC FlyBase: FBgn0003862; trx.
CC InterPro: IPR003889; FYRIC_H.
CC InterPro: IPR003888; FYRIC_N.
CC InterPro: IPR003616; PostSET.
CC InterPro: IPR001214; SET.
CC InterPro: IPR001965; ZnF_PHD.
CC InterPro: IPR001841; ZnF_ring.
CC Pfam: PF00628; PHD; 3.
CC Pfam: PF00856; SET; 1.
CC SMART: SM00542; FYRC; 1.

```


DR SMART: SM00541: EYRN. 1.
DR SMART: SM00249: PHD. 4.
DR SMART: SM00508: POSTSET. 1.
DR SMART: SM00184: RING. 3.
DR SMART: SM00317: SET. 1.
DR PROSITE: PS01280: SET. 1.
DR PROSITE: PS01359: ZF-PHD_1. 4.
DR PROSITE: PS0016: ZF-PHD_2. 3.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KW Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3599 3708 SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-SER.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VASPLC 1 368 MISSING (IN SHORT ISOFORM).
FT VASPLC 2025 2025 P -> PWLTSPKFLGSLTHGGLMLLLGVVRLKQGG
FT CONFLICT 2025 2025 (IN REF. 1).
FT CONFLICT 2341 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 2392 G -> S (IN REF. 1).
FT SEQUENCE 3726 AA; 400575 MW; D275650763D1CF5 CRC64;
Query Match 3.2%; Score 476; DB 1; Length 3726;
Best Local Similarity 18.8%; Pred. No. 1.1e-09;
Matches 628; Conservative 419; Mismatches 1138; Indels 1154; Gaps 155;
QY 5 EEEEDGAEETQSEED- EDEMEEDDDSDYPREMEDDDASYCTESSFRSHST 60
DB 538 DEDDEGVTFRRNDSPEDONNAEDDDDDDEDEEDDDONEDNDEASAKSAETKS 597
QY 61 YSS- TPGR- RKPRVHRPSLTLEEDIPPLEPKSSEDLMPN 101
DB 598 AGADRRDEKOLVDSHFVLPKRSTRSSRIKPKRLLEBGA- STKKRLSGD 651
QY 102 EHMNVIAI- YEVLNFGT- VLRLSPFRPE 129
DB 652 SKKRVFGTSSSSAGSTASTFSAITNLKKEFPNFGTIKNSAAGFVLRQRLQFQ 711
QY 130 DFCAALVSOECTLMAEHNVLKAVLREEDTSNTTGPADIKDSVNSTLYFIDGMTWE 189
DB 712 -ADNOQATFAPACPTSPSAIKPANSLATSSFGSLATNSSTV- TPTPS 760
QY 190 VLRYVCE- SDKEVHNHLYPOEADDPYGPVENKI- KVLDFLYDQFLTTIAEELMS 244
DB 761 ACST- CSAVVSKEY- TOARKYGVACDVCKRFFSKMKTKKSIASNSTAN- TS 810
QY 245 EGVLOYDHCHEVCHKLDGLL- CETCSAVYHLECVPLLEVPEDMEQCEVC- 295
DB 811 SGSGQY- LQCKGNCSGSPCSIHSAKSQJLKNPKFKFKD- RCTACWL 852
QY 296 -VAHK- VPGVT- DCVAEIQKNKPYIRHEPIGIDRSRRKYVFLNRR 337
DB 853 KKMISFOLPAHRSRLSAILPGRMGRGAARREKSAELL- SPTGSLRF- 900
QY 338 LIIEEDFENEKKIWIYSTKVOALALDCLDKQWEA- ELCKLLEMEBEIHHMIOIT 395
DB 901 -TSTASSSPSV- VASTVK- WKSSGSTALTSLIKPNPLAENNVY 943
QY 396 EDLTNKARGS- NKSFL- AAANEELIESIRAK- KKDINNVKSPEETEKDK 441
DB 944 FGSFPLRLPALLENLFLKISNADQKLAALAEALISPLTLKNSKQEKKEKYESESEKLL 1003

QY 442 NETENDSKD- AEKNREFEEDOSLEKSDDKTPDDDPQKSE- EPTVEVD 489
DB 1004 SPTQAGTKKSGAAEAQVEVPOPEKEAPQSTTTPPSANGASHGVPQALAGENATGD 1063
QY 490 KONSVANLGDNTVATSEETSPSEGRSPVGLSE- TPDSNMAEKKVASLEPODV- 544
DB 1064 TLKRRIDLPGRVHVCHVCASAVILQ- PLATFGEQDQPEDADMQEITAAVPSAIMP 1122
QY 545 -PEENKTCESSTNTSATTTSTIOPNLNENSSSELNSSQESAKAADDPENGERESHTPVS 603
DB 1123 SPEKPLTHIYTDENDNCASCSTSPVGDESKPSKSGSAQAEVKA- TA 1168
QY 604 IOBEIYDGTSEKSTGELSESPAGKAGSAGSTRITRLRNPNPSKLSQKSOVAAAAHHA 663
DB 1169 LCKE- GTASAGGSSAKVY- TKN- AAVAS 1194
QY 664 NKLPEGRKVLVNVSGEISRLSTKKEVIMKNNYFRLGQEGRYVYHNOYTNISFAL 723
DB 1195 NLIVASKK- QRNGDI- ATSSVTQSSN- OTQGRKYKEHROORT- L 1236
QY 724 NKHOREHDH- KRHLA- HKFCLT 745
DB 1237 ISIDWENDPRAVQOTGRGLIVETVAGRALCPLGSGTGLDPLIFCACCCEPHQYQVQ 1296
QY 746 PAGEFRKWSVHS- KYLTISTRLITTOLENNIPSEFPHNASHIRANNT- 795
DB 1297 DEYNLK- HGFEDDTLMGSLLETTVN- ASTGPSSTL- NOTLORLMLICRPTVC 1347
QY 796 -KAVQKCSK- PREFALALILBCAVPYVMLPIW 827
DB 1348 YTCNMSSGSKVQCKQKQKHYHSTCLGTSKRLLIGADRLICVCLCKSKSSTTKV 1401
QY 828 REFLG- HTRL- 836
DB 1402 SKFVGNLPMWCTGCFLLRKKNFCPIQRCYDNDNDLKMMEGCGDGQWHSCEGLSDEQ 1461
QY 837 HMTSTIEREEKKVKKKEKKEOEELTMOQATWYKTFPYKHQVYKQK- GEERYVTGYGW 895
DB 1462 YNLSTLPESIEFICKKARNSS- KIKAEWROAAVEEERKASLYVL 1509
QY 896 SWISTHYRFPYKLPGNINNVYRKSLBEGTKNNMENDESDSKRCSRPKIKTEP- 952
DB 1510 KILSKROACALLKLSPRKNV- KCTGASSNOGKIQDPRAL 1548
QY 953 -DSEKDEVKSDAKGADQDQNMDSIKITEKKQDVKELDSDSPCKEPEMEYDDDMK 1010
DB 1549 QFSSGSDNGLSD- GEGQNSDDVYEFKDDQOQ- QOQRNANMKP- RYKPL- PCS 1598
QY 1011 TESHVNCQES- SQVDVNVVSEGFHLRTSYKKTKSKSLDGLERRIKOFTLEEKQREXI 1069
DB 1599 COOHISHSQSFSLVDI- KOKIAGNSYVSLLEBNYDMSQVIOO- 1639
QY 1070 KLEGGIKGIGKSTNSKNLSPEYITKAKEGCQSDMKQESPANNDQ- PED 1122
DB 1640 -SNCDELIDIAVK- ELISEQFPWFQNETACTDALEED 1674
QY 1123 LIQCS- QSDSVILRMSDPSHTTNKLYPKDRLVDVDSISPETKCPKQNSIEND 1175
DB 1675 MFESGSGNGYEDLDQAGVSAVYNEHSTQAESGVID- IPL- 1717
QY 1176 IEEKVSDIASRGOEPTKSKTKGNDFIDSKLASADDIGTLCKNK- KPLI 1225
DB 1718 -EEVDDFGSGCIG- MRDITRMLCFCKRSGEGLSGEEARLILCGHDWCWHTNCAMSAEV 1774
QY 1226 QEESDTIYSSSKSALHSSPKS- TNDRODTPLSRADFEGKIGCQDES- N 1273
DB 1775 FEEDID- GSIQNVHSAVARGMTKCTVCGKRGAT- VGCNRYSGGEHYH 1819
QY 1274 STLENSDVTYSIODSS- EEDMIYONSNSISQOFTREQDVYLEPLCEIYS 1325
DB 1820 YPCASIDICAPLITDMSMCPAPAKGNALKNAGSPSV- TYISNFEVSRPYVEL- 1872
QY 1326 GESTGNCEDRLPVKGTGANGRKPSQOK- KLEER- PVN- KCS 1363


```

Db 1873 -----DRKKRLIE-----PARVOFHIGSLEVRQLGAIIPRSDSYEAVVPIFLCS 1919
Qy 1364 -----DOIKLKNITDKKNENRESE-----KQOFTSFQJNGKDNKRKILYKGCLE 1412
Db 1920 RLWSSKEPMKIEVEYTRTITQNSSSTLALDVGRNYIYDHTNPNPNSKEVQLMAQIARWH 1979
Qy 1413 IESER---VSGNVE-----PKVNNINKIIPENDIKSLVKESSAIPFENGIVIMEDFNE 1464
Db 1980 TSLARSEFLJENGTDWMSGFEPNNSC--VPPDONTBEEQOQADLIPRLKALIFEDL--- 2035
Qy 1465 RNSSEKSHLSSDAEGN---YRDSLETLPSKES--DSQTOTTPSASCPESSNVQVED 1520
Db 2036 -----PHELDGISMDLFIYDOKTDLFAISEQSKDGTQAMTS-----NOAQN 2078
Qy 1521 MELEISEVKVYSSPTTSEESNLMDIFDENGPI--KNNENVNGSKRKITYITEVTTMT 1579
Db 2079 QNOQAGANSVS---ICEDTRNSMTSLG--NGWPSANFVEEDAMLSAEN---SSQVWL 2130
Qy 1580 STVA-----TESKTVIYEKGDKOTVVSSTENCAKSTVTTTITVTIKLSTPGSYD 1632
Db 2131 KTLAMPKLDGNSAMATIKRRRLSK-----NLAEVFLILSSQORKKKEKATYAGV- 2181
Qy 1633 IISVKEOSTVVTYVTDLSLTGGTLVSM--VSKEYSTRD-----KVKLKEESP 1683
Db 2182 -----SRROSISETSVGEVATTSGSVRSKFTWSAKRYFEKSEGREEAKRMIMQMDG- 2235
Qy 1684 KTRRSTALPSYRKPFYKTSKSIPLPMDLKLARKGIR---EVPYFNNAFPAID 1739
Db 2236 -----VDDSIETFRITISGDNLSSTAOFSQOVKCDRCQCYRINDAOR-- 2278
Qy 1740 IMPYSPRPFTGLTWRYRQOTVKSAGVSLMLRLMLASLRMDMAKVPBGGSSTPETS 1799
Db 2279 --HLEPSCPT-----MSSNETSDVS-----GGMTNNANO 2307
Qy 1800 EHEITTEILKRRDVGPIREYICRKLICPIGVPEPKETPOR-----KGLR 1850
Db 2308 ISSESINELOKULANAGLINV-----LOSATSPFOVRLGSLGFGIQLQ 2354
Qy 1851 SSALRPK-----RPETPKOTGPVITETWAELEELW--EIRAFERV----- 1891
Db 2355 QILOQFOSLGNGFFLSQPNPATOANT-----DDLOITYANSLQGLAANLGGFFL 2403
Qy 1892 -----EKEKAQAVEQAKKRLQOKPVTIATSTTSPTSTSTISPAQKVAVAPI 1941
Db 2404 AOPTVAPAPOLIAVSTNPBGQCIQLIPQIMQATTPPTATYQULQAINTKKIML--- 2460
Qy 1942 SGSVTTGTAMVL--TKVGSPTATVTFQOKNRFHQFPATVK-----QGSNSGVVOVOO- 1993
Db 2461 --PLTAAGRPKLTVAATKAQAQAVKOROLKSGHOVKPIQAKLOPHQHQOQOQTOVQOP 2518
Qy 1994 -FVLG-----IIPSTGT-----SQQTFST----- 2013
Db 2519 ITVMGONLLOPOLLFOSSITQTOAPQIILPQAOPONIISVTDGSGOGPLOYITISPTAGE 2578
Qy 2014 ---QPR-TATVTRPNTSGSG-----TTSNS--QVITGPQIIRPGMTVIR 2052
Db 2579 YKBPQPTATPFLTPATPAGATYIQTDAAGNLVLTFTPSNSGLQMLTQOSIQAOQOVIG 2638
Qy 2053 TPLQOSTL-----GKAIITRPVWQCAPQOVM-----TOITNG 2086
Db 2639 TLIQPTIOLGGGADN-----QPSNOQPLLIGTGSSGLEFATTPSVYILAT 2689
Qy 2087 QPV-----STAVSAPNTVSSST-PG---OKSLTATSTNSIIOSSAOPRPPQGOYK 2133
Db 2690 QPMYIGLEITIVQITVWSSQOEVSTAMPGLSONASFATTTVOFQSKIEP-----IYD 2743
Qy 2134 LTMAQLTLOLTOGHGNOGLTV---VIOGOCO--TTGOLILPOGVTVLPBGQOLMQAA 2187
Db 2744 LAGYVVLNNTGDASAGFLNAASVLOQOQTDODDTTQI----- 2782
Qy 2188 MNGTVORFLPLPLATTATTTTSTTAAGTGEOROSKLSPOMOYHODKTLPPAOS 2247

```

```

Db 2783 LONANFO--FQSVPTSSGASTSMIDYTPVAVTA-----KIPYTOIKRTNAQAKAAG 2832
Qy 2248 --SVSGPAAKAOQ-----TAQBSAPROPOQPOSPAPQEVQTOPEVQOQTIVSSHVSE 2299
Db 2833 ISGVGVPPPOPVNVKNLPTSLVITQSOVQVKNNSLNKOSQVKGKAASGTTGCGA--PSSI 2891
Qy 2300 A-----QPTHAQSSKPOV-----AASQOPSNOVQSGPVRV--Q 2331
Db 2892 ASKPIQOKTNMTMRPIHKLEVKPKVMKPTPKYQONQNSHLLQOQOQOQPOLOQOQIIPAVVWQ 2951
Qy 2332 SP-----SOTRIRPST-----PSQLSPGOQSOVQTTTSOPT---PIQHTISQ 2371
Db 2952 VPKVTISQORIPAOIQOQOQLOQAOMIHIPOQOQOPILOQOQOVQVPSMPTITLAEAPVQSO 3011
Qy 2372 IPSQOQPOSOPOVQOSTORTLSSQOTLNOVSQSPSR----- 2407
Db 3012 FWEPOALFQOELANVQHFSTSSSSSSSSSSSSSLPTNVVPMQOQAPSTISSSTRPTNRV 3071
Qy 2408 -POLQIQOPOP-----QVAVNPOLQOQVO--VLQIOQOVNAQIOAQOQSGVPOQIKLOLP 2459
Db 3072 LPMQORQEPAPLISNECPVYSSPTPKPVPBQPIIHQMTSASVSKCYAQKSTLPSV---YE 3128
Qy 2460 IQLOQSSAVQTHQIONVTVQOASVQOUL-----QVQOQLRDQOQKK 2502
Db 3129 AELKVSVALS--IYDVTMDALIEQPYQSYTEGLYKENSPEGSKTEOLLLOQOQRE 3186
Qy 2503 O--QOQIENVNTPSKLLIVEILQOVYMKHNAVIEH-----LKOKKS--MTPARREN 2552
Db 3187 QLNQOLVNNGYLLDKHTFOVEPRMDTVYREEDLEEBEDDEDFSLKMRISACNDHEMSDS 3246
Qy 2553 QRMIVQOVMYILDKIDKE-----KQAAKKR 2581
Db 3247 EEPVAVDKISK--ILDNLTDNDCADSIATATMEVDASAGYQQWVEDVLTATTAQASAPTEE 3305
Qy 2582 REBSVQKRSKONATYLSLKRKQOLAELIKRALLD--KDLQLEQBELKRLDKTK 2640
Db 3306 FEGALETAAVEAATYINMADAN-----YIDLKQLOLNGVELLRR--KE 3349
Qy 2641 EKDLMOAOTAVAAACPPVTPVLPAPAPPPSPPPPG 2679
Db 3350 EORTVSQEOBQSKALVP--TAAAPRPPPIQEPKMTG 3386

RESULT 3
K167_HUMAN
ID K167_HUMAN STANDARD: PRI: 3256 AA.
AC P46013;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen KI-67.
GN MKI67.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043435; PubMed=8227122;
RA Schlueter C., Duchrow M., Mohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody KI-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL
CC PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1
CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO

```


QY	1402	KIYLKGECLKEISSRKYSGV-EPKVYNNIKLILPENDKLSLTVESAIRPPTINDVIME	1460
Db	1431	INAFRETAJAKOLDBAASVYTSKRHHPTKR--EKAKPLEDLAGW--KELFOTP-----VCT	1480
QY	1461	DENERNSESTKSHLSSSDAGCNVROSLLETLPSTKESDSTOTFTTPSASCPEBNSVNOVED	1520
Db	1481	DKPPTHETTKIACROSPD-----PVDFTPSKPOSKRBLKRYD	1519
QY	1521	MEIETSEVKKYVTS-----PTSEBSNLSNDFIDENGLPLNK--NENVNGESKR	1568
Db	1520	VEEEFFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM--GTPOKDLDTLENLGSKR	1574
QY	1569	KTVILEYTTMYSYA-----TSKTYIKYKEDKQTVVSNEMACKSVTTTT	1616
Db	1575	LOTREKRAQALEDLAGEKELFOTRGHTEBSM--NDKTAACKSSOPDLDKNASSKR	1631
QY	1617	TYVYKLSPTSGYSDVILSYEOKSKVTVTYVYVDSLTTTGVLVSMYVSEKSTRDYVK	1676
Db	1632	RLKTSLG--KVGYKEELLAVGKILQTSGETTHHTEPTGDGSKMAFMSPQILDSAS	1669
QY	1677	LMKSSRKRTRSJA-----LPSTRK--FYTKSTKSIVLPNDL-----	1715
Db	1690	LTSKRQRLRPEKCESEVEDLAGEIELFOTPSHTKSMTEKTTVYSAPODLVDVPT	1749
QY	1716	-----KKIARGGIRVVPYFNNAKPALDINWYPSRPFTGIRWRYLQYVKSILAGVSL	1769
Db	1750	SSKPOKPSRLKAPTEE-----EFLAFKQTSAG-----KAMHPKPAVGEK	1793
QY	1770	MLR-LMAWSLRMDMAAKVPBGGSSTRETSETTEITTEITIKRDVPGYIR-----	1820
Db	1794	DINFILCTPOVKLDOPGNLF--GSNRLQOTREKKAQALELT-----GFELEFOTPCT	1844
QY	1821	-----FECYIKIIC--PIGV--EPPEKPTPOKRLBS--ALRPRRP-----	1859
Db	1845	DNLPADKEKTKIKLCKSPQSDPADDTPTNKKOPKRSKLKADVEEFLAKRLTPSAGAM	1904
QY	1901	QOAKRLEQOKPYVIAVSTSPSTTSITSPISPAQKMAVAPISGVYTGKMWLTKVQSP	1960
Db	1905	HTPKAA-----VGESEK--DINTFQVPEKLDLLGNJPGSKRRPQPTREKAKALE	1952
QY	1953	DLAGEKELFOTRGHTEBSMDDKITVEYCKSPDPDPKTP-----TSKKORLKSICK-	2005
QY	1961	ATVFOONKNFHQFAFWWAKOGSNGSVVOQOKVL--GIIPSTGTSQOTFTSPORTA	2018
Db	2006	-----YGVKEVLPVGLKLTQSGKTTQT-----	2028
QY	2019	TYVIRPNTSGSGT-----TSNSQVLT--TGPOIRGMVYIRP-----LQOSTLG	2061
Db	2029	-----HRETAGDGSKIKAFKESASOMLDPANYGTGMGRWP-----TRPKEMASLEDLAGE	2079
QY	2062	KAIRITVWVOPGAPQOVMOIIRGQVPSTAVSAPVTSFPGOKLSIATSIHQSSA	2121
Db	2080	KELQOTDHNHESTTDDKTKI-----AKSPPESMOTPIST-----	2117
QY	2122	SQPPPOOGGVKLT--MAOLLTOLGHHGNGLTVVIOGOGTQOLOIP--OGVTV	2175
Db	2118	RRRKRTPLKGRDIVEELSALKOLOT-----TNHTKVPDEDEKGIN	2159
QY	2176	LPGHGOOLM-----QAMPMGYVO-----RFLP--PLATATTAATSTTTTV	2215
Db	2160	FRETAJAKOLDBAASVYTSKQOPTRPKAOPRLEDLAGLKELEFOTVCAJDKPLTHEKTTKI	2219
QY	2216	STTA-----AGTEB--OROSKLS--POMOVHODKTLPPAOSSSVGAQAKOPOTA	2260
Db	2220	ACRSPOCPDVGTPPIEFKQPSKRSRLKADVEBSLAKRKTBSVGRKAMOTPKPAGDEKDM	2279
QY	2261	-----QPSARPOQOTPOSPAPREVOOTOPVOTOTTVSS-----HVSESAOPT	2303
Db	2280	KAFMGTVQKRLDPLGNLPSGKRMPQPKR-----KQALEDLAGEKELFOTRGHTEBTPCT	2332

QY	2304	HAOSSKPGVAAOSPOGNNVCCOSGSPVNRVQSSQRIKPS-----	-TSS-	2344
		::: ::		
Db	2333	-TDEKTKRIACKS-PQ-----	-PDPVDPASTKQRPKNLKKADVEEFLAKRRTPSA	2383
QY	2345	-----QLSPGQSOOVQWTTSQPIQPIQHNLSL-QIP-SQGQPOSQP-----		2382
QY	2383	-QVQSSQTQTLSSGQTLQNVSVSSPSRPQIQIQPQPIVAVP-----	QIQQOVQLSQI-	2435
Db	2444	FQTGHTHEESMTDKITVEYSCSPQESPESFTSSKQRIKIPLVKYMKEEPLAVSKLTR	250303	
QY	2436	QSQVAVQIQAOQSSVPQIKL-QLPQIQOQSSAVQTHQIQNVVYQQAASVQEQIQIQOQ	2493	
Db	2504	TSGETQTQHTPEPTGDSKSIKAFKESPQIIDLPA-----	-SVTSRRQQLRTREKARA	2555
QY	2494	LRDQOQKK-----QOQIEINVN-----	-PSKLLIVQELIQ-----	-KQVYKH 2531
Db	2556	LEDVADFEKEFLSAPGHTHEESMTDKNKIKIPCKSPPELTDITSTRKCPKTRPKKEVKEE	2615	
QY	2532	NAVIEHLKQKKSMT-----PAEREENQRMVQNMVYKILDKIDKEQQAKKKREES	2585	
Db	2616	LSAVERLQTQSGSTHTHKRPAAGDEGIKVL-----	-KQRAKKKPNVE	2658
QY	2586	VEQKRSKQNA-----TKISALIFKFKQLRA-----	-ELKKRA	2618
Db	2659	EEPSRRRRPAPKEAQLIEDLAGFETELSETSGHTQESTLAGKATKIPCESPFLVEVDTA	2718	
QY	2619	LDDQDQIQEVOEELKRLDKIKKKDKMLQIAQ-----	-ATAV	2653
Db	2719	STRRLRLRRVQK-----VQYKEEPSAKKFTQISGETTDADKEPAGEDKGIKALKESAQT	2773	
QY	2654	AAPCPVTPVLPAPAPPPSP-----PPPGVQHTGLTLSTPLPVASQKRKREEKD	2705	
Db	2774	PAPASATVSGSRRRAPAREQAQIEDLAGKQDPAQAHTEESMTDKTIKICKSSPELED	2833	
QY	2706	SSSKSSKKKKMISTTSKTKKQDKLYLCIKTPYDESKFYIGCDRCQNMVHGRCVGILOSEA	2765	
Db	2834	TATSSKRRPRTRAQKVEKEE-LLAIVGK-----	-LTQSG	2866
QY	2766	ELIDEVCPQCGSTEDAMVLPLETRKQVDEGLKRLVRLSQAHKAMWPLEPVDN-TPAD	2824	
Db	2867	E-----THTHTRKEVGBEGKTKA-----	-TKQPAKRVNDAED	2897
QY	2825	YVGIKEPM-----DLATMEERVQR-RYEKTEFEVD	2856	
Db	2898	VIGSRQRPAPRAKKAQPLDELASQELSTQPHTEBLAQAAD	2940	
RESULT 4				
ANK2_HUMAN				
ID	ANK2_HUMAN	STANDARD;	PRT;	3924 AA.
AC	001484; 001485;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, noneythrhold).			
GN	ANK2.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).			
RC	TISSUE=Brain stem;			
FX	MEDLINE=91302466; PubMed=1830053;			
RA	Otto F., Kunimoto M., McLaughlin T., Bennett V.;			
RT	"Isolation and characterization of cDNAs encoding human brain			
RL	ankyrins reveal a family of alternatively spliced genes.";			
	J. Cell Biol. 114:241-255(1991).			
	[2]			
RP	REVISIONS.			
RA	Carpenter S.;			

Db 1778 PSLKSEBHAAGSPSPKTERHSTLSSAKTERBHPVPSPSKTEKHSPSPSAKTERHSPAS 1837
QY 616 KSTGELISEFGAGKAGSGSTRITRLRNDPSKLSQKSOVAAAHAANKLREKGEVLV 675
Db 1838 SSSKTEKHSPP-----VSPSTRK--TERHSPVSTKTERBHPVSPSG-----KTDKRPPV 1883
QY 676 VNSOGELISRTKKEVIMGNINNYFKLQEGKYRVYHNQYSTNSPALNKHQHREDHDK- 734
Db 1884 -----SPSGTEKHPPVSEKTEKRLVPSPSGR-----TKHQQVSTAGCT 1924
QY 735 RRLHAHKFCLTPAGEFKMWSVHGSVLTJLTLRLTTOLENNIPSSFFHPMAHSHRANV 794
Db 1925 EKHLP-----VSPSGTEKOPVSP-----TSKTERIEETMSVREL----- 1960
QY 795 IKAVQMSKREFALALALIECAVKPVMLPIREFLGHRLHMRMSIEREEKYK-KK 853
Db 1961 MKAFOGQODPSKHKTGTFEHSKAKOK-----OPOEGKAVREK 1998
QY 854 EK-----KOEETMOQATVWKYTFPVKHOVMKQGEYKVTGGWSWISKTHVRYEVP 908
Db 1999 EKGPIILQREAOCTENOTI-----KRGORLPTVG----- 2027
QY 909 KLPGNTNVTNRKSLGKTKNNMDESDKRCRSRSPKTIKTPDSEKDEKGSDAKGA 968
Db 2028 --TAESKRGVRSISGVKK--EDAAAGKREKVLSHKIPEVQSVPEBESHRESEVPREKMA 2083
QY 969 D-QNEMDISKITEKK--DOVKELIDSDDKPCKEEPEVDDM-KTESHVNCOESSOVDP 1025
Db 2084 DEQGDMDLQISPDKRTSTDESVYIKQELENDKYOQFRLSEETEKQOHLHD-----QVLT 2138
QY 1026 VNVSEGFHLRTSYKK-----RTKSSKLDLLE-----RIKQFTLEF 1062
Db 2139 SPFWTTPPL--DYMKDEFLPALSIQSGALDGSSELSKNEGAVSPGSLMEGTPOLISSE 2196
QY 1063 KORLEKIKLGGIKGIGKICTNSKNLSSEPVITKAKEGQSDSMQEOJPNANNO-- 1119
Db 2197 SYKHE-----GLAETPETSPELSFSPKSEBOTGETKSTKTEYTTETRSEKHP 2247
QY 1120 -PEDLIOGCSQSDSVYRMSDPSHTTKLYPKRVLLDVVIRSPEKCPKONSLEND-- 1175
Db 2248 TTKKITGSGSERGATV--TEDSETSPESFOKEATLG--SPKDISPRK--QDDCTG 2296
QY 1176 -----IEEKVSDLASR--GQEPYKSKTKGNDFFLDDSKLASADDTIGTLICK 1219
Db 2297 SCVALAKETPTGLTEBAACDEGORTGSSAHKTQT-----DSEAOQS-----TATSD 2344
QY 1220 NKRKLIOEESDTIVSSSKSLHSSVSPKSTNDROTPLSRAMDFEGKICGCSSESSTLENS 1279
Db 2345 ETKALPLPEASVKTDTGTESRPOGAVIRSPQGLELALPSR-----DSEVLASAVAD 2394
QY 1280 SDTVSIODSSEEDMIVONSNESISEOPRTREODVEVLEPLKCELVSGESTGNCEDRL-- 1336
Db 2395 SLAASHDLSLEASPVLEDNNS-----HKTPDLSLEP-----SPLKSPCNDLSLESS 2439
QY 1337 PYKGTENGKPKPS-----OQKULEERVANKCSD 1364
Db 2440 PVEPKMAGITFPHPLPAVAKTELLTEVAVSARLLRDPDGAEDSDLEQTSIMESSG 2499
QY 1365 QIKLKNTTDKNNNRESEKKGQRTSTFOJNGKNDKPKYILKGCELEKIESRVSNGVE 1424
Db 2500 KSPILSPPTPSEEVSYEVTKTTVST-----PKRAVHHCACEDD-----SNGE 2545
QY 1425 PKVNNINKIIPENDIKSLTVKESAIRPFLNGDVIMEDFNERNSSETSKSHLSSDAEGNY 1484
Db 2546 KK-----RFPPEEMFMKVTIKIMFDELEQAKOKRDYKREPKOESS-----SSSDPDADC 2597
QY 1485 RDSLETLPSTRESSTOTTTTPSACSPESNSVNOVEDMEIETSEVKKYS----- 1533
Db 2598 SVDVDE--PKHTGSEDESGVPLVLTSESRRVSSSESEPELAQLKKADSGLLPEPIRV 2656
QY 1534 ---SPITSEESNLSNFI-----DENGILPINKENUNNGES-----KRKT 1570
Db 2657 QPSPPLPSSMDSNSPPEVOPPVYVSKOYTTKMNEDTQOEBFGKSEEEKDSSHLAEDRHA 2716

QY 1571 VITEVTTMTSTVATESKFIKVEKGDQKVYSTENCAKSTVTTTTVTKLSPSTGGS 1630
Db 2717 VSTE-----ADRSYDKLNROTDQPKIODGHGCEMSPSSSARVSSGLOSPGDD 2767
QY 1631 VDIISVKEQS--KVVVTTTTVDSLTGTLVTSMT-----VSKESTRDKKYLKMF 1680
Db 2768 VD-----EQPVLYKESIALOGTHEKDEGEELDVRAESPQADCPSESSESSSLPHCLV 2822
QY 1681 SRPKTRSG--TALPSYRKRVTSKKSIFLFPNDLKLARKGGLREVYPFYNKAPALD 1739
Db 2823 SEGKLEDEDISATSSIOKTEVTKTDBTENLPK----- 2856
QY 1740 IMPYSPRPPTFGITWYRLOQTVKSLAGVSLMLRLMLASLRWDMAMAKVPPGGSTRTETS 1799
Db 2857 -----CPQSDSSIT--TQT-----DRFSMDVPVSDLAENDEY 2887
QY 1800 EFERITTEIIRKRDNDGPPYIREYCIKRLICPIGV-----ETPKETPPQKGLRS 1851
Db 2888 DPOLITS-----PYE-----NVPSSQSFSSSESKTQTDANHTTSEHS 2923
QY 1852 SALRPKRPEP-----KOTGPVLIETWAAEELMELMIRAFAREYKEKAQAVEQQA 1903
Db 2924 SEYVSTTITSPVEDVYVASSSGTVLSK-----ESNPEGDIK-----MESGL 2966
QY 1904 KKRLL--BQKRPYIATSTTSPTSTSTISPAOKVMVAPISGVTTGTCKWYL--TVGSPA 1961
Db 2967 ESTLWEMQSDSV--SSSFPTMSATYTVGEOI-----SKVYIKTKTVDSD- 3010
QY 1962 TYTFQONKMFHQTFATVWVQGSNSGVVOQKVLGILPSSTGTSOQRTFSQPTATVVT 2021
Db 3011 --SWSEIRDEDEAFEARV-----BEOKITGLM-----VDRQ----- 3041
QY 2022 IRPNTSGSGTTSNSQVITGPQIRPGMTVIRTPLOOSTLGAALIRTPVWVQAPQOVVT 2081
Db 3042 -----SGQTT-----PDTTPARPTEEGT--PTSEQNPFLFGQSK-----LF 3076
QY 2082 QITRGQPVSTASAPNTVS--STPGOKSLTATSTISNIQSASQPPRQOQOVULTMAQ 2138
Db 3077 EMTRSGAIDMTKRSYADESHFPOIOESREETLSDVKEGATGADPLLE-----TSAE 3131
QY 2139 LQJLQGHGNGOGLJVYVIGOGQGTGOLQIPIGVV--VLPFGQOQLMAAMPGTVO 2194
Db 3132 SLALS-----BSKETVDEADLLIPDSVSEVEYELPASAQULN----- 3168
QY 2195 RFLFTPLATATATSTTTTIVSTTAAGTGEOROSKLSPOMVQHODKTLPPAQSSSVGPAP 2254
Db 3169 -----SQMGISASTETPTKCAVSVT-----KDLPTVQGTGDIPLS 3204
QY 2255 AOPQTAQ--SARQOPQO-----POSPAP-----EVQOQPEVQQT 2290
Db 3205 GVRQISCPDPSSEPAVOVDFTLTRSVMYSDRGDSDPSSEPOKSVIETPTAPMENVPF 3264
QY 2291 TVS--SHVPSAOP-----HAQSKP-----QVA 2313
Db 3265 TESKSKIPVMTPTSTIPAPSAYESVSDEFTLSSVDEENKADKPKSLPVKVPVLQAV 3324
QY 2314 AQSOPQSNVQGSPPVRVQSPQRTIRPSPQSLSPPQOQSOVOTTSQPIPIOPHTSLQIP 2373
Db 3325 EQGLSDLDITSVOKTVAPQGDMAASIAPDNRK--SESDASSLSKTKCPKVTBRSYETETE 3383
QY 2374 SOQOPQ-----SOQOVOSTQTLSSGQVLNQV--SVSSPSRPOLO----- 2411
Db 3384 SREARELESESEGAIRPILITSLRPLVNSRSTSSCRGOTSTPKSKKEHFDLYNSTIE 3443
QY 2412 -----IOQPOQYIAYVQLOQOVVL.SQIQSQVYVAQIOAQSGVPOQIKLOLPIQIQSS 2466
Db 3444 FFEESIDPASKLYDRLOQSREQEIVYSDDESSSALAVSYENLPPVETESHVVEDIFDR 3503
QY 2467 AVQTHQIONVVT-----VQAAVQEOLOQVQ-----QLRQOQOKKQ 2503
Db 3504 PIMDESITETLIERLPDENGHDHAEPPQDEQERIEERLAYIADHLGFSWTELARELDFTEE 3563

QY 2504 QQEIIVNPNPSKL-----LIKV-----EITOKOVYMKHNAVIEHL 2538
DB 3564 QHQIIEIPNPNISQDOYQLKIMLERDGNHATDNLVECLTINKIMDYLHMETWTEPL 3623
QY 2539 KOKKSTPAERENQOMITONQMYKILD-----KIDKEKQAKKKRRESVEOKRS-- 2591
DB 3624 QERISHSYAIEBOT-----ITLHSGEFSYLOEBLCTAOKHKEGQAVSSESET 3672
QY 2592 -----KONATKLSALLF--KHKEQLRAEILKKRALLDKDLQ 2625
DB 3673 CDHPPIVSEEDISVGYSTFQDGVKTEGDSSTALFPQTHKEGVQDQFSGKMDLPEESS 3732
QY 2626 IEVQELAKDLKIKKEDLMQLAQATAVAAPCPVP-----VLPAPPAPEPPPPPGV 2680
DB 3733 LEVQDEY-----FVTTPTGETSETKAMIVPSPSPKTEPEVSTPAE 3773
QY 2681 QHTGLSTPT-----LPVASOKRKEEKKDSSSKKKMMISTSKETKDKDKLCTCK 2734
DB 3774 EKKLYIQTPTSSRGGSPILIOPEEPSEHREESS--PRKTSLVESADNQPET-----CE 3827
QY 2735 TPYDESKFYIGCDRCOMWYHRCVGILOSEBELIDEVQPOCOSTEDAMTLPLTEK-- 2792
DB 3828 RLDEDAFAFKGDMP-----IPPEVYEEYI-----DEHGHTVYAKVTARKII 3871
QY 2793 ----DYEGLRKVLRSLOAHKMAWPLEVPDNDAPDYGVK 2830
DB 3872 RRYVSEGETEKEIINVQGMPO-----EPVNIIEGDSYKVIK 3908

RESULT 5

TRX_DROVI STANDARD; PRT; 3828 AA.

AC 024742;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Trithorax protein.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrochi L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; 250038; CAA90349.1; -
DR HSSP; P20393; 1A6Y.
DR Flybase; FBgn0014844; Dvir\trx.
DR InterPro; IPR003889; Fyricl.C.
DR InterPro; IPR003888; Fyricl.N.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.

DR InterPro; IPR001628; Znf_Casteroid.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 3.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00508; PostSET; 1.
DR SMART; SM00184; RING; 2.
DR SMART; SM00317; SET; 1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS01359; ZF_PHD_1; 3.
DR PROSITE; PS50016; ZF_PHD_2; 3.
KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator.
FT ZN_FING 1251 1334 PHD-type 1.
FT ZN_FING 1335 1380 PHD-type 2.
FT ZN_FING 1408 1469 PHD-type 3.
FT ZN_FING 1708 1767 PHD-type 4 (ATYPICAL).
FT ZN_FING 1768 1818 PHD-type 5 (ATYPICAL).
FT DOMAIN 3701 3810 SET.
FT DOMAIN 28 41 POLY-ALA.
FT DOMAIN 66 71 POLY-ASP.
FT DOMAIN 160 164 POLY-ASP.
FT DOMAIN 173 182 POLY-ALA.
FT DOMAIN 221 228 POLY-GLN.
FT DOMAIN 243 251 POLY-ALA.
FT DOMAIN 253 258 POLY-THR.
FT DOMAIN 292 296 POLY-ALA.
FT DOMAIN 538 546 POLY-ASP.
FT DOMAIN 1072 1075 POLY-GLU.
FT DOMAIN 2483 3271 GLN-RICH.
FT DOMAIN 3333 3339 POLY-ASP.
SQ SEQUENCE 3828 AA; 413721 MW; 32059CF303A3C504 CRC64;

Query Match 2.7%; Score 407.5; DB 1; Length 3828;
Best local similarity 18.6%; Pred. No. 3e-07;
Matches 691; Conservative 461; Mismatches 1260; Indels 1315; Gaps 177;

QY 5 EEEEDGDAEETDSDDEDEDEEDDDSDVPEEM-----EDDDDDASYCT 51
DB 523 EDNDDGSA-----SSDALEDEDIDDDAEEENEFAASEKSAETASVDEKEADQLVM 577
QY 52 ESSE--RSHSTYSTGRRKPRVHRPSPILEKDIPLPEFPSSSDLMVPRHI-----M 105
DB 578 DKHFVLPKRSTRSS-----RIIKPNKRILEVGGICSKRSPSDANGKPKKYPGLATL 630
QY 106 NVIAIEVLNFGTVL--RLSPRFEDFCALVSQOCFLMAEMHYLLKAVLREEDTSN 163
DB 631 PACCTRRRRSATATLSQKRGKTFASFATKYNSS-----FVLNQPRQFQTDKS 681
QY 164 TTFGPADLDSVNSTLYFLDGMTPVRLRYCESDKYHHVLP-----YQEAEDPYGPV 218
DB 682 RSEVSA--KPTLPPTT-----VLPASSAITSANVLSFGAL 715
QY 219 ENKIKVLQPLVDFLTNTNAREELMSEGVYQYDDHRCVC-----HKGLDGLCC 266
DB 716 NN-----ANSAAVAASTGAVCSAPVNNKADPLARKYG--VIAC 751
QY 267 ETC-----SAVYHLECVKPPLEEYPE-----DEMOCEVCAVHVGVDCAVEIQKRP 315
DB 752 EYVKRKNRKRTKSKLSTPHNSPSTSTASQGLKC-----TD----- 790
QY 316 YIRHEPIGYDRSRKKYWFNLRLIIEEDTENENEKKIWIYSTIKVQLAELIDCLDKDYWA 375
DB 791 -----GGNCS-----ILSLKSQLKNFKKL--YKERCKACWMLKCIATLQLPA 830
QY 376 -----ELCKTL--EEMREIIRHMDITEDL-----TNKAR 403
DB 831 GHRSLRSLAILPASMREVAAPKDKCPDLLSPATSLRFTAFTSSASSGTTIKWSSAETAV 890

QY 404 GSKSFLAANE-----ELLESTAKGIDIDNVKSPEETE----- 438
 Db 891 NSIKSNPLANNYTFGGTPLLRAILLEPLFLIGSDNKKAKESKEKLEGLSPVSTSEAA 950
 QY 439 -KXKNFENDSKAEKNREFEEDOSLEKSDDKTPDDPEQGSSEB---TEVGDKNSV 494
 Db 951 VADGKTRKAKOKREKARELEAKPLSPNAKKTTEANTPETQKDEOPASPTTTVSAASS 1010
 QY 495 SANIGDNTNATSEETSPSEGRS----- 517
 Db 1011 TSHTSAAATNSQLETTTEANANASAVPDNLKROIDLGPVKKHVCSASIVLGOPLATFG 1070
 QY 518 -----PVCLISETPDSSNNAEKKVAVSELPODVEEPNKTCESSNTSAT---TTS 563
 Db 1071 DEEBELAAEAGAPPTTTTTSPEVILIKRKPSPQWQMTIDENDCASCILPTTEATAE 1130
 QY 564 IOBNLEN-SNSSLSEINSSOSSEKAAADDPENGRESHTPVSIOEIVGDTSEKSGELS 622
 Db 1131 AOPAVSVLESKSSKSTQTEARK-----TPA-----TSGSSGKVT 1167
 QY 623 ESPGAKGASGSTRITIRLNPDKLSQLSQOAAAAHAHANKLFEGKEVILVYNSQEL 682
 Db 1168 TRNATFTVTSVASSIVATKORNIIEVSSSISSQAAATOSRRALAK-----EV 1215
 QY 683 SRLSTKEVIMKGNINNY-----FKLOEGK-----Y 709
 Db 1216 NRL---KALLSIDFWENYDPAEYCOTGFLITVEYAOARALCFELCSTGIDPLIFCACC 1272
 QY 710 RYVHNOYSTNSFALNKHOREDHDKRHLAKCLTPAGFEKNGSVHSGKVLITSLRL 769
 Db 1273 EPHYQCVLDEYML-KHSEED-----TLMT 1297
 QY 770 TITOLENNITPSPFHPNWA---SHRANWI-----KAYOMCSK----- 803
 Db 1298 SLETSNNACAIASAANTALNQLTORLNMLCPRCTVOCYCNMSSGKVKQKQKXVHST 1357
 QY 804 -----PREFALALILECAVKVYVMLPWIREPLGTRILHRMTSIREKEKY 860
 Db 1358 CIGTSKRLLGADRLICVNLCKCKSCATTKV-----SKEVGN-LPMCTAC-----FKL 1404
 QY 851 KKK-----EKKOEBEE---TMOQA---TWY-----KY---TFPV----- 875
 Db 1405 RKGNGNCPICQKCYDNDPDLKMECGDCNQWHSKCEGISDEQYMLLSTLPISIEFICK 1464
 QY 876 -----KHQVKK-KGEEYRVYGVGWSMISTKHYRFPKLPQNTNANRSKLE 923
 Db 1465 KCARCQDVSNKADENRQAWMEEFKSIVYLKLSKRSQACALLL---SPKNNMCCSA 1522
 QY 924 GTRNNNDENWDESDKRCSSPKIKITIEPDESEDEVKGSADAKGADONEMDISKITEKED 983
 Db 1523 GAOP-----AKAHSQKLOPKALOFTYNGL-GSDGESQNSDDIYEFEKQ-- 1565
 QY 984 QDVKELLSDSDCKCKEPEPEVDDMKTESHVNCQSSQVDVYVNBEGFLIRTSYKKTR 1043
 Db 1566 -----HSTNKRFPSTPVP-----CSCLOPLSQSPSFL-VDIKOKTA 1600
 QY 1044 SSKLDGLLEIRIKOFTLEEKORLEKIKLEGGIKIGIKTSTNSKNLSESEVITRAKEGCO 1103
 Db 1601 SNAYVSLAE-----FNVDMSQVIQO-----SNODELDIAK----- 1631
 QY 1104 SDSMKROESFNANNDQ-----PEDLIGCSQSDSSVLRMSDPSTTKLYPK---DR 1152
 Db 1632 --ELLSEQFPWFQNETKACTDALEEDMFESCGYEELKESPTTYAHHHTASQARTGLDII 1689
 QY 1153 VLDUVSIRBPETKCPKONSIENDIEEKVDSLARQGEPLTKSKTK-----GNDFIDSKL 1207
 Db 1690 PLDDVDVDLG---GC---AVKTRLDTRVCLFCRKSSEGLSGEABRLLYGCHDQVH--- 1738
 QY 1208 ASADIDIGTLICKRKRPLIOESPTIVSSSKSALHSVPKS-----TMDROAT----- 1254
 Db 1739 -----INCAMMSAEVEEID---GSLQVNHSAVANGRMKICVCGKRGATVCCNWK 1786
 QY 1255 -----PLSRAMD---FEGRIGCDESENSTLE-NSSDTVSIOSSSEEDMIVQNSNES 1301

Db 1787 SCGEHYHPCARDIDAFILDKSMYCPAHARNALKANGSPVSYESNFEV-----SRPV 1840
 QY 1302 ISBQFTRBEDVAVLEPLKCEL-----VSGSENCEDR-----LPVGTBA-----NO 1345
 Db 1841 YVELERKR---KLIVPAKQPHIGSVAVNROLSIYPRPSDSEALVPIFLCSRLYWS 1897
 QY 1346 KRPSQOKKLEERP-----NKCSDQIKL---KNTRDKNNENRESERKQ----- 1386
 Db 1898 KEP---WKIYEIVTRTIQSYSSSTLILADGRNFYTDHNTPNCSLVLQLAQIARHSSL 1954
 QY 1387 QRTSTQJONGKDNKPKIYKGECLKEISRSRVYSGVNERVNNINKIIPENDIKSLTYE 1446
 Db 1955 ARSDLLDTWAE-FPNSYVPAD---ENTIE---EPQ-QVADLILPEE--IKD-AIRE 1999
 QY 1447 SAIRPFLNG---DVIM-EDFNERN-----SSEPKSHLLSSSAEGNYDSLETLPSFKE 1496
 Db 2000 DLPHELLIDGISMDIDFMYEELGDKTELPAWSEOSKGTJATSOAG--ASVILCDEDTN 2057
 QY 1497 SDS-TQTTTPSASCPESSNVQY-----EDMEIFTSYV-KKVTSSP-----I 1536
 Db 2058 SNLKNHVLVSNCTIASNPVDDAMLCAARSSQSEKCGDVLKKTDTAPTRSMWKLDGGSV 2117
 QY 1537 TSEESNLNDFTDENGLPJKNKNENNGESKRTVTEVTYMTSVATESKYIYKEKD 1596
 Db 2118 AAFKRRRLSKINIAE--GYLLSLNO---RSKK-----EMATVAGITR-- 2153
 QY 1597 KOTVSTENCASIVYTTTITVTKLSTPSTGGSVDIISYKESK-----FVVTIT-VT 1649
 Db 2154 KOSVCGSSELPAEGSATMRTKSP-----TWSACLEFKENESREPAKLTIMONDGDV 2206
 QY 1650 DSLTT---TGGLTVTSMTVSK-----EXSTRDKVK-----LMKESRPKRTS 1688
 Db 2207 DSITEYRIIGSDQNLSTAQGTGVKCEQOCYRANVDSQRHLHGCSEPMSTSESEFAT 2266
 QY 1689 GTALBPYRKFTVTKSTKSLFVLPNDLKK-----LARKGIGREPYRYNNAKPLDII 1741
 Db 2267 GTAQLSAESL-----NELQOALAAATLSMNGL-----NY---LQTSF 2302
 QY 1742 PYSPRPPT---FGITWRYRLQTVKSLAGVL-----MURLMASL 1778
 Db 2303 POYQNLATLIGQFVQVQLOGLQTLQ--LOPOSIGNGFLSOPMAQATNSGNDVQLATANS 2361
 QY 1779 RMDMAKVPYGGGSTRTESETEITTEILIKRNDVPGYIREYCIRKILCPIGVETP 1838
 Db 2362 Q-NLANL--GGGFLTLQPTMSTQAPOLI-ALSTNPGS-----TQ 2398
 QY 1839 KETPTPQKGLNSSLAR---PKRPETPKOTGPVILIEFWAEELEIMETRAFEREYERK 1895
 Db 2399 OFIOLFOGNGATTOLOLQTAAPLRCNATYOT---LOATNSDKKIVLFLFEA-GDPLQEVY 2453
 QY 1896 AOAVEO-----QAKRLEBQOKTVIATSTPSTSTSTISPAOKV 1936
 Db 2454 TOAAQOATAAHOKOLKSGHVKPQIAKLOGQOQOORHOQHOQOQOQOQOQOQOQO 2513
 QY 1937 MWAPISGYVTGTKNVLTLLTKVGSPPATVTEQONKNFQOTFAWYKQOSGNSGVYQOKVL 1996
 Db 2514 QOPPIVAAQHGGTTOQLGQNLQOP--QLFQSNMQ--PQOQOLLIPQOQAN----- 2561
 QY 1997 GIIPSSTGTSQO-----TFTSPORPATVITIRPNTSGSGGT----- 2033
 Db 2562 -LISFVTGGGSONQLOVYISIPPTTNDFKPQOQTTSPFTFLTAGGATFLQTDASGNMLLT 2620
 QY 2034 -----SNSQVITGPQIRPBMVYTRTPLOQSTL---GKALITRPVWVP----- 2073
 Db 2621 TAPANSGLQMLTG-QLOTOPQVYIGTLIQPOTLQTLTGADGTAPAAQOPLLIGATGGST 2679
 QY 2074 -----GAPOQVWTOITIRQOPV-----STAVASAPTVESST-PG---OKSLASATST 2114
 Db 2680 TGLEPATAQOVIILAT---QPMYIGLETLIVONTVWSSQOPFVSTAMPBGVLSQNSSSATIT 2735
 QY 2115 SNTQSSASOP---PRP-----OQGVKLTMAOULTQJLQGHGNGOGLT 2153


```

Db 2736 QVQASKEIPYVLPAGYVVLNNAVDASGNTSLQSOQTATDATALQQLQ-NAGFOFOT 2794
QY 2154 VVIQGGQOTTGOLQIIPQVGT-----VLPSPGQ-OLMQAA 2187
Db 2795 TPTTSTQQTSTDYAPRLVYAKVPVQAOKRMTNANKSPISVLSKVQRPQOSQVYVNV 2854
QY 2188 MPMGTV-----QRFLETPLATTAATTATTTT-----VSTTAG--T 2222
Db 2855 LPTNVVIOQQQQQQQQQQQQQQQQQQQQQLLGNMNLKLTISOFORQQQANLKNQAAGQQT 2914
QY 2223 GEGROSKLSPQMVHDKT--LPPAOSSSYGR--AKAQPOTAPRSARPOQOTQPSRA-- 2276
Db 2915 GSTGCAAPSLASKPLQKTKMNLIRPIHKEVEYKPKIMQAPKLASASWQHQQQSPAAI 2974
QY 2277 -----QPEVQTPQVQOTQTVSSHNPSAPQPTHAOSKPYAAQSQPQ 2319
Db 2975 NQVAKVALLQRLAPARQRP-QQEPQEQHNLHQQQQQQQQQQQHQHQHQQQQQLSMFQ 3033
QY 2320 SNVQGSPPV-----RVQSPQSTRIRPSTRPQSLSPQSQSOVQTTTSQPIR----- 2363
Db 3034 L-LRAQQPPIISIVNTAEPQATQFVIRPALQAOAPITQIQEQSQQQQQQPAEQILINGKA 3092
QY 2364 --IQPHTSLQIPSQ-QQPOQRPQVOSTOTLSSGQTLNQSVSQSPR-----POLQIQ 2414
Db 3093 ARLQRYASNSLPTNVVNPLOQQRCASANNSSNVYQONSTTINSRPTNRVLPMQQRQ 3152
QY 2415 PGP-----QVIAVP-----QLQQQVQVLSQSQVY---AQIOAQQSGVPOQIKLQLP1-Q1 2462
Db 3153 PPLMSNDVVVQSPPTPRKPIEPYRPAGASTQKPIVKCYAQLEQKSPGQETLKNITLIDL 3212
QY 2463 QQSSAVQTHQIQ-----NVYVQOASVQEQLOL-----VQQLRQOQKRRQ 2505
Db 3213 EQNTSITTMQLOQPOQGPPIGEQIFEQSEBAQVQLEKPKHMDLMLLEATSCQOQQQQQOH 3272
QY 2506 IETNVMTPSKLILKVELIQOVYMKHNAVIE-----HLKQKS-----MT 2545
Db 3273 MEVNVNGEQLVSN-----ESCLLEKGFNVYEAUPMDETHYASKMGGSGGAEGIGQVD 3328
QY 2546 PAREEN-----QRMTCN-----QVMKYLIDKIDKEE--KOAKRRR 2582
Db 3329 DAEDEDDDDDFSLKMATSNACNDHMSDSEPAVKKEISILDNLTNDDSDSLATATTV 3388
QY 2583 ESEVEQRKSKQN--ATKLSALFKHKEQLRA-----ETIKRALLDKDLQIE 2627
Db 3389 EASAGYQOMVEDVLTATAGSVSTDETFATAEAVEAASYNEMAEHQLKQLOAG 3448
QY 2628 VQBELRDLKIKKEXDLMQATAVAAQPCRPVLPAPAPRPPSPRPVGHGILS 2687
Db 3449 VELDLK-----KPKLDVPOQOQPTVP--PNVVPYAAAPQPPRMDPKK-----IS 3492
QY 2688 TPTLPVASQKRREBEKDSKSKKKMISTSKETKDKTKLCYCKTPDES----- 2740
Db 3493 GPHLLKEIQSEDOFTYKSSSLAEIWEKVFPAVYVARAHG-----LTPLEPGLADMSG 3546
QY 2741 -----KTYI-----GCDRCQNMW-----HGRCVG-----ILQSE 2764
Db 3547 VQWIGLKTNAKYLIEQLGVEKCVKTYPKYHRRNGNVSTAAGGAHRTGSPAPALAG 3606
QY 2765 AE-LID-----EYVCPOQOSTEDAMVLTPLTEKDYGLKRVLSLOAHKMAWP 2812
Db 3607 AELIDVYGSDDQELQENAEACARCE-----PYVSSEYDMFSWLSRRHKOPTQV 3656
QY 2813 FLEVPVNDAP-----DYGVVIEKPM-----DL----- 2835
Db 3657 FYQPSNENLVPRGTSNLPJMAWKYRIKETYDYGVFRSHHGGLYCTKRIEAGEMY 3716
QY 2836 -----ATMEERVQRRYVE-----KITE--FVADMKIFPNCRYNDSDP 2873
Db 3717 IEYAGELLIRSTLTDK--RERYDSDRGIGCYMFKIDMLVVDATWRGAARFINHSCBP 3772

```

RESULT 6
ATRX_HUMAN

```

ID ATRX_HUMAN STANDARD; PRT; 2492 AA.
AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
DE nuclear protein) (XNP) (Znf-HX).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
RP VARIANTS ATR-X.
RX MEDLINE=97123494; PubMed=8968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RX MEDLINE=97386582; PubMed=9244431;
RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
RA Colleaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 860-2492 FROM N.A.
RX MEDLINE=95179111; PubMed=874112;
RA Strayton C.L., Dabovic B., Gulsano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Strayton C.L.,
RA Villasseau P., Khrestchatskiy M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP SEQUENCE OF 1375-2492 FROM N.A.
RA Pearce A., Chapman J.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP EZH2 BINDING.
RX MEDLINE=98167853; PubMed=9499421;
RA Cardoso C., Timsit S., Villard L., Khrestchatskiy M., Fontes M.,
RA Colleaux L.;
RT "Specific interaction between the XNP/ATRX gene product and the SET
RT domain of the human EZH2 protein.";
RL Hum. Mol. Genet. 7:679-684(1998).
RN [8]
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
RP HETEROCHROMATIN.
RX MEDLINE=20040663; PubMed=10570185;
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
RT "Localization of a putative transcriptional regulator (ATRX) at

```


Db 161 GEDGLHGT-----VSTACGQGVNH---FQKDSITRHPSLQ--VLICKNCEKYYM 205
 QY 234 TTNIABEELMSEGVIOYDCHCRVCHKLGLDJLCCETCSAVYHLCV-----KPLEEVED 288
 Db 206 SDDISND---SDGM-----DECCRCWACBGNLICDPCFNACCKICLRLNKLSTIMDE 259
 QY 289 --EMOCEVCAHVP---GVTDG--VAE-----IQKNKPYIRHEPIGYDRSRKRYWELNR 336
 Db 260 NQWMYCIC--HPEPLDLVTACNSVFNLEQLQONKKIK---VDSEKSNKYEHSTR 314
 QY 337 RLIEEDTE--NENEKKI-----WYSTKQVLAELIDCLDKDWEALCLILEMRE 366
 Db 315 FSPKKTSSNCGEKKLDSCSGSVTYSALLVPKEM-----KRAKKLLETTAN 365
 QY 387 EIRHMDITEDLTNKAQSN-----KSFLAANEELIESIRAKKDDIIN--YKSP 435
 Db 366 MNSYVFLKQATDINSISSATKLROLAKFVLA-----DIKKAHLALEDLIDSEFPAMD 421
 QY 436 ETEKDKNETEN---DSKDAKNREEDFOSLEKSDDKT-----PDDDEQKSEEPTE 486
 Db 422 AVNKEKNTKEHKYIDAKFETKARKGEKPCALEKKIDISKSPAKLSRKQVDESHMHQNVPT 481
 QY 487 VQKGNVSANLNDNTNATSEETSPSEGR--SPVGCLETFPDDSSNAEKVASELPQDVP 545
 Db 482 -----BORTKSTGGEHKKSDDRKEEPQEPANTESELDMDIVSPSSVPEDIF 529
 QY 546 EEBNKTC-----SSNTSATTTSTIOPNLNENSSSETSOSSESAKAADPENECERE 597
 Db 530 ENETAMEVOSSVDHOGDSSGTBOEVESSVYKLNISKNRGGIKSKITA---KYTKE 565
 QY 598 SH---TPVSTIQ-----EELVGFSTSEKSTG--ELSESPAGKAGASSTRIITR-- 640
 Db 586 LYLKLPVSLPNSPIGADQOEVPQDKDYGKSCGLNPKLEKCGHGNDSNEHJVENEVS 645
 QY 641 -----LNPBSKLSQLK-----SQOVAANHEANKLEKEGEVLYV-----NSQ 679
 Db 646 LLIEESDLRSPVKPTTLPRLPETNEVYTSNDEECNEIYKOKLSVPRKKDKRNSD 705
 QY 680 GEISR-----LSTKKEVIMKGNINNYFKLQEGKYRVH-----NOYSTNS---F 721
 Db 706 SALDNKRPKNLPRKSKOSEIYDQNSDSEMLAILKGYSRMHSSSSDPDIDIEHTNHTLY 765
 QY 722 ALKKHOREDHDKRHLAHKFCFLTPAGEF--KNGSVHGSKVLITSLRLTLTQLENNIP 779
 Db 766 DLTOAGKDKGRKRRKSS---TSGSDFTKKGRKSAKS---IISKRRQROSESS-- 815
 QY 780 SSFFHPRMASHRANMIKAYOMCSKPREPALALILECAVVPVMLPIWREFLGHTLRHM 839
 Db 816 -----NYSELEKEIKSMKIGAAR-----TTKKRLIPNT 844
 QY 840 TSIEREKEKVKKK-----EKKOEETMOQATWVKYTFPVKHQ 878
 Db 845 KDDSDSEDEKSHKKGMDNQHKLKTSQESSSDAEKQORE-----TF----- 888
 QY 879 VMKQGEERYVNTGYGSMWISKTHVYRFPKLPNGNTNVNRYKSLGKTKNMMDENDES 938
 Db 889 -----SSAEGTVDDOTIMELRD- 906
 QY 939 RKCSRPKTIKIPDESKDEYKSDAAKQADQ--EMDISKITEKKQD-----VKEL 989
 Db 907 ---RLPKQOQASATD---GVDLKSGKQSFSLTEVRKVAETLKREKSHLTKTKCKY 957
 QY 990 LQSDSDPK---CKEPEMEYDDMKTESHYNCOESSQVDDVNVNSEGHLTKSYKK--- 1041
 Db 958 QDLSLDAEFLKDKQSDSEISDDKQSKKGTTEKKK-----PSDRKKVYIKM 1005
 QY 1042 ---TKSKLIDGLER-----RIKQFTLEKQRLERIKLEGGIK----- 1076
 Db 1006 EQQYESSDDTEKLPREHEICHFPKGIKQIKNGTJDEKSKSKIRKTSKKDELSDYAE 1065
 QY 1077 ---GIGKTSSNKNLSESVYITKAKEGCO--SDSMQOESPPANNDQPEDLIQGCQSD 1131
 Db 1066 KSTGKGSDSCSDSKSKNGAYGREKKRCKLKGSSRRKRD-----CSSSD 1111

QY 1132 SYVLKMSDPSTHTNKLXPDRVLVDVYSIRSPETKCPKONSIE-----NDIEKVSOLA 1184
 Db 1112 TEKYSMEDECCNS-----DKRLKRIELRERRNLSSKRNKTEILOGSSSSDAESSDDNK 1166
 QY 1185 SRGOEPTKSKTKGNDFFIDSKLASADDITGLICKNKKPLIOESDTIYSSKSALHSSV 1244
 Db 1167 KKKQR--TSSKKK-----AVIYKEK-----RNSLSTST 1193
 QY 1245 PKSTNDATPLBRAMDFEKLQGDSESNSTLENSSTVSIOD-----SSEEDMIVON 1297
 Db 1194 KKKQADITTS---SSSSDIE-----DDQNSIGESSDQKIKPYTENLVLSHFGCQSS 1245
 QY 1298 SNESISEQFRTQDYVE-----VLBPCLKELVSGESTGNCEDRLPYKTEAN 1344
 Db 1246 GDEALSKSVAVTYDDDDDDNDPENRIAKKMLLEIKANLSSDDE--GSSDDE--PEEGKRT 1303
 QY 1345 GKRPQOKKLEB--RPVNCSDQ-----IKLKNTT-----DKKNNEBRE 1381
 Db 1304 GKONEENPGDEEAKNOVNSSESDSESKKPRYRHRLLRKLTVSDGSGEKKTKPREH 1363
 QY 1382 SEKKQRTSTFOJNGKONKPKIKYLGECLEKISESRVYSGNVPEKVVNINKIIPENDIKS 1441
 Db 1364 KEVKGRRNR--KVSSSESDSDFOEGSVSEVSBS---EDQRPRTSARKALEENORS 1418
 QY 1442 LTVKESAIRPFIINGDYIMEDFNERNSSETKSHLSSDAGNRYDSLETLPSTRES---- 1497
 Db 1419 YKQKKRRRIKQVEDSSSE--NKSNSDEEKEKEEKEEEDDEEDNDOSKSPGKG 1476
 QY 1498 -----DSTQTTTPSACPEBSNVQVEDMEIETSEVKVY-----TSSPITS- 1538
 Db 1477 RKRIRKLKDKDKLRTETONALKEEERRRKRIAEERERERKLEIEDASPTKPIYTK 1536
 QY 1539 -----BEESNLNDIFDENLPIKNKNENVG-----ESKRKT-----VITEVT 1576
 Db 1537 LVUDEDETEKPLVOYRRNVIKIKPHQVGVOPMDCCESVAKTKKSPSGCIIAHCM 1596
 QY 1577 TMTSTVATES--KTVIKERKQKVY 1601
 Db 1597 GLGKTLQVSEFLHTVLLCDKLDPESTAL 1623

RESULT 7
 ANK3_HUMAN
 ID ANK3_HUMAN STANDARD; PRT: 4377 AA.
 AC 012955;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ankyrin 3 (ANK-3) (Ankyrin G).
 GN ANK3.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain stem;
 RX MEDLINE=9518209; PubMed=7836469;
 RA Kordeli E., Lambert S., Bennett V.,
 RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized
 at the axonal initial segment and node of Ranvier.";
 RJ J. Biol. Chem. 270:2352-2359(1995).
 CC -!- FUNCTION: Membrane-cytoskeleton linker.
 CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by
 CC alternative splicing.
 CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues.
 CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
 CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U13616; AAA64834.1; -
DR HSSP: P55273; 1B18.
DR Genew: HGNC:494; ANK3.
DR MIM: 600465; -
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000488; Death.
DR InterPro: IPR000906; ZU5.
DR Pfam: PF00023; ank; 24.
DR Pfam: PF00531; death; 1.
DR Pfam: PF00791; ZU5; 1.
DR PRINTS: PR01415; ANKRYN.
DR SMART: SM00248; ANK; 21.
DR SMART: SM00005; DEATH; 1.
DR SMART: SM00218; ZU5; 1.
DR PROSITE: PS50088; ANK_REPEAT; 21.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat.
FT REPEAT 73 102 ANK 1.
FT REPEAT 106 135 ANK 2.
FT REPEAT 139 168 ANK 3.
FT REPEAT 172 201 ANK 4.
FT REPEAT 203 230 ANK 5.
FT REPEAT 234 263 ANK 6.
FT REPEAT 267 296 ANK 7.
FT REPEAT 300 329 ANK 8.
FT REPEAT 333 362 ANK 9.
FT REPEAT 366 395 ANK 10.
FT REPEAT 399 428 ANK 11.
FT REPEAT 432 461 ANK 12.
FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT REPEAT 825 854 ANK 24.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA; 480399 MW; FA2379E55768B684 CRC64.

Query Match 2.78; Score 406; DB 1; Length 4377;
Best Local Similarity 17.68; Pred. No. 3.9e-07;
Matches 558; Conservative 444; Mismatches 1118; Indels 1048; Gaps 126;

QY 17 QDSEDDDEDEDEDDDDSDY-----PEEMEDDDDDA----- 47
DB 1441 KETSSDDDELKTRKQASALAKRKRYLLEPGMTERSTGATNSLPPTYYSKPFST 1500
QY 48 ----SYCT-----ESSFRSHSTYSSTPGRRKP-----RVHRPRSPILLEKDIPLLE 89
DB 1501 RPYQSMTTADITVGPAPKSGFTSLSSSSSNTMPSASPLKSIWVSTP-SPIKSTIGASTTS 1559
QY 90 FPRSSSEDLMPNHEIMVIAIYEVLRNFGVLRISPPREFDCAALYSQECYTLMAEMHV 149
DB 1560 SVYSISIVASP-----INSLRTMSSPIKTVVSSQSPYNIQVSSGTLARAPAVTEAPFLKG 1613
QY 150 VLLKAVLRREDTSNTTGG-----PAULKDSVNSTLYFIDQMTPEVLAVYCES 197
DB 1614 LAANSSTFSSSTSVTTAGSLIERSSTITMTPPAPSKNTIN----- 1653
QY 198 DKLEYHNLVPO-----EAEDYRPGVVENKIKVLQFLVDOFLFTNIMAEELMSEGVIOYDHD 253
DB 1654 ---YSSSLPFRKSLITSAAPLISSPLKSVSPVAKSRVDVSSAKITMASSLSISSVYKQMPGH 1710

QY 254 CRVCHKLGLD-----LCCECSAVYHLECVAPLEEVPEDEWQCEVCAHKVPQVT 304
DB 1711 AEVALVNSISPLKXASSSTLINGCKATATLD-----EKSSATNSVSSVSAT 1760
QY 305 DCVAETQKN-----KPYTRHEPGYDNRRR-----YWFNLRLIIEEDTENENEK 350
DB 1761 DTVEKVFSTTAMPFSPRLSRVYSAPSAFOSLRTPSASALYTSLGSSISATTSVTSII 1820
QY 351 KIMVYST-----KVQALIDICLDVWEALCKLLEEMREIIN--RIMDITTEDLTNKR 403
DB 1821 TVPYSVVNVNLEPALKLPPNSFTKSAALISPIKTLTTEHPQPHFSRTSPV---- 1876
QY 404 GSNKSFLLA-----ANEEILIESIRAKKDIDNVKS-----PEETE 438
DB 1877 -KSLFLPALPALKLISTPSSLSQOELKDVAMKEDLMRMALITQDVPEKPPQPELPK 1935
QY 439 KDKNETENDSKDAEKNREFEDQS--LEKD--SDDKTPDDPEQKSEPTVEGDKGNSV 494
DB 1936 EGRIDDEEPFKIVEKVEDLVKVSSEILKKDYCVDNK--GSPKSPKS-----DKGHP 1985
QY 495 SANIGDNTTNTSEETSPSEGRSPVGCSETPDSSNMAEKVASELIPQDVPEEN-KTCE 553
DB 1986 E-----DDWTFSSSEI--REAR-----QDAASQSP-SLEPVRQVAKA 2022
QY 554 SSNTSATTTSIQPLNENSSSELNSQSESAKAADPENERESHPP-VSTQEE----- 607
DB 2023 ASEKDYNLTQYIDLNTDNGSSSLNLKRYKPEDAKKQEGQKVLPAILOHEHKLMP 2082
QY 608 --IVGDTSEKSTGBLSESPGAKGASGSTRITRLRNPD-----SKLS---QLK 652
DB 2083 PASMRSTSEKEELCKMADS-----FETDTRI--LESPPDQFQHDQKSPLSDSGFE 2133
QY 653 SQOYAAAHAE-----KLEKE-----GKEVLV-----NSGELSRLSTKKEVI 692
DB 2134 SEKTPSPASQAEETGPKPLFHEVPPIPVYETRETVVHNRSTYPSGVDVPOQPEEPVS 2193
QY 693 MKGNINNYFKLGQGYRVYHNOYSTNSFALNKHQHREDHDKRRHLAKKCLTPAGEFKW 752
DB 2194 PKPS-PPEMEL--EPKPTTSIKKVKAFQMKASSEDDHN--RVLS----- 2255
QY 753 NGSHGSKVLITSLRLTLTQLENNIPSSFFPHWASHRA-----NWKAVQMSKPR 805
DB 2236 ---KGMRYKEETHTITTRVYVHSPPG--EGASERIEETMSVDIDKAFSGRDP 2287
QY 806 EFALALILBCAYKPVVMLPIWREFLHTRLHRM--TSTIEREKEVKKKKEKQEEET 862
DB 2288 KELAGLEEHKSAVSPDY-----HKSAAETSAQHAEKKQMKPKLERITE-- 2331
QY 863 MQQATWVKYTFPVYKHQYWKQGEERYVTGYGQMSWISKTIVYRFPKLPQNTVNVFKSL 922
DB 2332 -----VHIEKGNQAEPTLV-----IIRTKNHPREKEMVUQKDL 2365
QY 923 -EGTKNMDEMDESKRKSRSPPKKIKIEPDSKDKDVYKSDAAKADQEMDISKTEK 961
DB 2366 SRGDINLKDLPEKHDFPQSE-----OGQOEEELTAEEESPILSYLSESRVNTVPSQ 2419
QY 982 KDO--DVKEILDSDQPKP---EEMPEVDD--MKTSVHNCQSSQVDVVNVSEGF 1032
DB 2420 EDSRPSAQLISDSTYTKLLLSQHSIEYHDELSELRGSGSYFAEKMLSEKLDV--- 2476
QY 1033 HLRTSYK---KTKSSKLDGLLER-RIKQFTLEKQRLKLEKIGIKIGTSTNSSK 1087
DB 2477 HSDTEESVTDHAPPSSELOGSDKRSREKIAIAPKKEIISKIYKDVSENGVAKVS----- 2531
QY 1088 NLSSEPVITAKKGCQSDSKROQSPANNQDEPDLIOGCSQSDSVLRMSDPSHT---- 1143
DB 2532 -----KDEHFDDVTVLHYSGN-----VSSPKHAMMMR 2558
QY 1144 -----TNKLYPKRDVLDVYSIRSPETKCKONSIMENDIEKYSDLASRQGEF-TKS 1193
DB 2559 FTEDRLDRGNEKLITIEDRV--DRTVKAEBEKLTEVSQFFRDKTEKLD--ELQSPKKA 2613


```

QY 1194 KTGNDFFIDSKLASAND---IGTLCKNKK-----PLIOEESDTIVSSS 1236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2614 RPNNGREYSQSPTSSPEKVLLELLASNDENVKARQHGPDGQFPKAEKAPSLPSSP 2673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1237 KSLHSSVPESTNDPRAT-----PLSRAMD-----FPGKLGCDs 1270
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2674 EKNVLS---QOTDSKSTYEAkakSISOKAPDQPSQFOLKQSKLSSIRLKFQGHAKS 2730
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1271 ESNSTLENSDVTs-----IDDS-----SEEDMIVQNSNISISPO-----FTRR 1309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2731 KDMGQDRKSDGOSRIIPVKKIQESKLPIYQVFAREKQKAKIDLDDSSVSQKQFMVLKTK 2790
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1310 EDOVEVLE-----PLKCELVSCESTGNCEDRL 1336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2791 DEHAQNEIYVANDSGDNKKQRTEMSKAMPDSFEOQAQKDLACHITSLATRGPMDDK 2850
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1337 PVKGTANG--KKPSOOKLKEERPANKSCDQIKLNTTQKKNNE--NRESEKQGRSTF 1392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2851 VFTWSSSGATNNKSOKEKLSHYLVH-----DVRENHIGHPESKSYDQKNEFM 2898
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1393 QINGKDNKPKIYLKGBCLKEISERSVSGNVEPKVNNINKIIPENDIKSLTVKESAIRP 1452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2899 SYTERERK---LITNGSLSEIKEMTVKSPS-----KKLYREIYVKEG 2938
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1453 INDGVI MEDPENNSSETKSH-----LISSDAEGNYSDELTPSTRESOSTOT 1502
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2939 DHGGLDPOPSRRSSSAVSHIPVRAVERRMILSNIPDQFCOS--APFKHELQKLSQ 2996
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1503 TTPSASCPSNSVNOVEDMEIETSEVKVP-----SSPTISE----- 1540
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2997 SSMKSETVETQHNSTIEDEKVTYSEISKVSKHOSYVGLCPLEETETSPKSPDLSLESP 3056
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1541 -EENLNDIFDENGRLPINKNNY-----NGESKRRTVITEVTTMTSTVATESKTVIKVE 1593
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3057 GKSPSSDVVDHS--PIDGLEKLAPLAQTEGKEIKTLPIYV--SFVGQVQYKEIEIQ 3110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1594 KGRKQVVSSTEMCAKSTYTTTITVTKLSTBSTGGSVDIISVKEOSKIVYTTTVDLSLT 1653
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3111 QGGVKKIISOECKVQETRTCTFTYTRQOKOPSPQSPEDDTLEQVS-----FLD 3160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1654 TTGGLVTSMT-----VSKYSTRDVKYKLMKF-----SRPKKTSSTALPS 1694
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3161 SSKSKSLTPBPSESEVSEYFETSKTPDSLAIYIPGKPSPIPEVSESESEEOAKSTSL- 3218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1695 YRKEVTKSTKKSIFV--LPNDLKLKLRKGGIREVYFENYNAKPAID----- 1739
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3219 ---KQTVYEETAVEREMNDVSKDNOCKNNRYATIEFPPLDADQIISDKKHNY 3273
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1740 -----IMVPS-----RPFGGITW 1754
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3274 LPEKEVDMIEVNLQDEHDKYQLAEPYIRVOP--PSVPBGADVSDDSEISYQPVVKKY 3332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1755 RYRLQTV-----KSLAGVSLMLRLIAMS-----LRDDMAAKYPPGGGSRTE 1797
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3333 TFFLKEVDEQKPKASAKASNOKELESNGKDNFGLGDSQNETAQNNDQST 3392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1798 TSETTTTTTIIIRRD-----VGPYGIREF-----YCIKRII 1829
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3393 TEGSINTTAFESHDTATEISLDGLDQDEDEDGLTESDSKLPQAMEIKKDIWNTEGIL 3452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1830 CFIG-----VPEPIETPTPQKGLRSSALRPKRPETPKQGVLIETVAEEL 1879
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3453 KPADRSFSQSKLEVIEEGEVGPDDEKPPKSSSSSEKTPDKTKGAGQFTLEGRRPDR 3512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1880 ELMEIAPFAERVEKEKAQAVEQA-----KKRLQOKPYIATS 1918
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3513 SVFPDTYFISKVEDEPATPKVATATGGLDFDPMSNNRGDEVDKSRDEETPRFGIAYE 3572
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1919 TTSPTSTSTISPAQKVMVAPISGVTGTGKWLTTKVGSPATVFOQKNNHOFATW 1978
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3573 DRSP--ATTPDITPAR-----TTPDESTPTSEPPFPFHCKMEEMT----- 3612
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1979 VKOGGNSGVVOYQKVLGIIPSTSTGSOQTFTSPQRTATVIRPNT-----SGSGGTTs 2034
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 3613 -----RSGAIDM-----SKRDFV--BERLOFQIGETSGKSGDQEGD 3650
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2035 NSOVIITGPMPVIRPMTYIRPLEOSTLGAIIIRTPVWVOPAPQOVMTQIIRGPVSTAVS 2094
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3651 KSMVATPQPOSQDITVETLERN-----VETPT-VEP-----NSIFPSGE 3691
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2095 APMTVSTPGQKSLTSATSTSNIOSSASOPPRPOGOVKLTMAQLTQLOGHGNGLTV 2154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3692 CQGTSSSGSLEKSAATNTSKVDPLRTP-----IKMGISAST-----MTM 3733
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2155 VIOGOGTTGQDOLIPQGVVLVPGGOQLMQAAMPKCTVORFLPLATTATTAATTTT 2214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3734 KKEGPEITDKIAY-----MTSCQGLENETITM 3762
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2215 VSTTAAGTGQROSKLSPQMVQHDKTLPPQAQSSVGAQAKQPTAPASRPOQ--TOP 2272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3763 ISNTA-----NSOMGVRFHEKIDFQK 3783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2273 QSPAQEPVOTQPEVQOTVSSHVSEAPOTHAOSSKPOVAAGSOPSNVQOSPVVQOS 2332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3784 DNFNNNNNLDSSITQTDNIMSNITVE-----HSAPTCCTEKDNPVKVS 3828
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2333 PSQTRI-----RPSTPSQLSPGOQOYQTTSPPIQP--HTSLQIPSQ 2375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3829 GKRTGVLOGCHVDKQKVLGEQOKTELIGIRQKSKLIRKATSPKDTFFPNHMSNTKASK 3888
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2376 GQPOPOVQOSSITQLSSGOTLNOVSSPSRPOQLQOQPOVIVPQLOQOVVLSQI 2435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3889 MKVOSSSEKTKALTTSSCDVVKSRIPKN-----TPRDNILAV----- 3926
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2436 QSQVNAQIQAO--QSGVPOQIKLQPIQIDQSSAVQTHQIDQNVVYQAAVQOBLQRYOOL 2494
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3927 RKACATQKQOQPKGAKQOLPKPVKVRSTCVTTTATTTATTTT 3975
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2495 RDOQOKKKQOQIENVTPEKLIKVEITQKQVYMKHNAVIEHLKOKKSTPAREBNOR 2554
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3976 -----TTTSCVTKVRSQLEKVCKHS--IEYFKG----- 4002
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2555 MIVCNOVMKYILDKIDKEEQAAKKRRREESVDOKSKONATSLALFKKQOLRA--- 2611
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4003 --ISGETLK-LVRLISEE---KKMOSELDEESTSRN-TLS-----ETSGGQ 4047
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2612 ELIKRALDKDQIQIEVQELKRLKIKKEK--DLQMQATQAVAAPC 2657
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4048 SVTTKSA---RDKKTEA-----APLKSSEKAKSEKSSRRTPQSPC 4087
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
AMNH_YEAST
ID AMNH_YEAST STANDARD: PRT: 1367 AA.
AC P08640: P08068:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase SI/s2 precursor (EC 3.2.1.3) (Glucaan 1,4-alpha-
glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN STAI OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrille B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churche C.M., Connor R., Copsey T., Dear S., Davila K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles S., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (Dec-1994) to the EMBL/GenBank/DBJ databases.
[2]

```


RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE-87194600; PubMed-3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN-SPX101-1C;
RX MEDLINE-89031230; PubMed-3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the StrA and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPCC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38061; CA86176.1; -;
DR EMBL: M16164; AA35014.1; -;
DR EMBL: M16165; AA35015.1; -;
DR EMBL: X13857; CA32069.1; -;
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001456; MUC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH
FT CARBOHYD 817 817 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 2.7%; Score 401.5; DB 1; Length 1367;
Best Local Similarity 21.2%; Pred. No. 1.5e-07;
Matches 289; Conservative 166; Mismatches 564; Indels 339; Gaps 56;

QY 1261 DFEQ---KLGQDSE-----SNSLTLENSDVTVSIDSSSEDMIVONSNEISIQFRT 1308
DB 191 DPEGFYWNIDCDNNGCGTKSTSTTSSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 243
QY 1309 REQVEVLEPLKCLVSESGNC--EDRLPVKGTENGKRPQOQKLEFPVKKCSQOI 1366
DB 244 TTSSTSESTSTSTAPAPPTTTSCTKRPPTTTSCTKRPPTTTSCTKRPPTTTSCTKRPPTT 296
QY 1367 KLTATTDKKNNESESEKGRSTFOJNCKDKPKIYLKCEKLEISESHVSGVNEPK 1426
DB 297 KKKTTTSTCTKTTTTPPTPSSSTTE---SSAPVPTPSSSTTESSAPVTSSTTE 350
QY 1427 VNNINKIIPENDIKSLVKEISAIRPFINGDIVIMEDFNERNSETEKSHLLSSDAEGNYRD 1486
DB 351 --SSAPVPTPSSSTTESSAPVTS-----STTESSAPVTSSTTESSAP 394
QY 1487 SLTELPSTKESDS---TQTTPASCEPSNSVNOVEDMEITSEVKKVTTSPTISE---- 1539
DB 395 VPTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTES 453
QY 1540 -----EESN--LSNDFIDENGLPIKKNENNVNGESKRRKTYTEVTMTST--VA 1583
DB 454 SSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV 513

QY 1584 TESKTVIKVEGDKQIVVSTSTENCASKSTVTTTTVTYTKLSTPSTGSGVDIISVKEQSKV 1643
DB 514 TPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSS-----STTESSSTP 568
QY 1644 VTTTVDLSLTGTTGTLVSMVSKKE--YSTRDYKLLKMRPKKTRSGTLPLSTRKTVK 1701
DB 569 VTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTS 628
QY 1702 STKRSIVLENDOLKILARKGRIREVPYFNKPKALDIMPSPRPTFCITMRRLQTV 1761
DB 629 STTESSA-----PVPPTSS-----TTE 647
QY 1762 KSLAGVSLMLRLMLASLRMDMAKVPGGGSTRTESETEITP-----TELIRK 1811
DB 648 SSAPVPTP-----SSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVTSSTTE 701
QY 1812 RDVGPYQIREYCIKRIKICPIGVN-----ETREKPTPORKGIRSSAL----- 1854
DB 702 SSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESS 761
QY 1855 ---RPRRPETPKOTGPVITETWVAEELEMEIRAFAEVEKEKAAVQOAKKRLREQ 1910
DB 762 APVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSNTTSS 821
QY 1911 KP--TVIATSTTS-----PTSTSTTSPAKQVMVAPISGVTTGCKMVLTTKVGSS-- 1959
DB 822 APSSTPSSSTTESSVPTPSSSTTESS-----APVSSSTTESSVAPVPTPSSSSN 874
QY 1960 ---PATVTFQOKNKNHOFATVYKQGSNGVYQOAKVLG-----IIPS----- 2001
DB 875 ITSSAPPSIIPFSSTTESFST-GTVTPSSSKYPSQOETSVSTTEVTIIVPTKTTSSVT 933
QY 2002 -----STGTSQ--OTFTSFOPRTAVTIRPMTSGSGTSSQVITSPQIRPM 2048
DB 934 PSTTTTITTCSTGTNSAGETTSQCSKTYTIVY-PTTTTVTSTSTTIT-----T 965
QY 2049 TVIRTPLO--QSTVLG--KAIIRT-PVMVQCAPQOVMQIIRGQVSTAVS-----AP 2096
DB 966 TVCSTGTNSAGETTSQCSKPTITTVTPCSTPSESTASPTSPPTVTVTVTVTTE 1045
QY 2097 NTVSSSTGQSLNSATSTNSIQSS--ASQPPRPQOQV-KLTMAQLQLQNGHNGUGLT 2153
DB 1046 YSTSTKSGGE-ITVTEVTKNIPVTVLTIAPTBPSVTVTVNPTTITTVTCSTGTNSA-- 1102
QY 2154 VVIQOGQTTGQLOLIPQGVTL---PCPQOQLQMAAMNGTVQRFPLATTAATAS 2209
DB 1103 -----GETTSGCS--PKVITTVPCSTGTGGEYTTTA-----TTIVTTA----- 1138
QY 2210 TTTTTVSTTAAGTGEOROSKLSPOMOVHOKTLEPPAOSSSVGPAPKAOPTAOPSARPQO 2269
DB 1139 -VTTTVTVTSTSTGTNSAGKTTTG--YTKKSVPTTVTVTLAPS----- 1178
QY 2270 TQPSAPQAPVQOPEVQOTVTVSHVPSAOPTHAOSKPOVAQOSQPOSNNVQOGSPVR 2329
DB 1179 ---APVTPATNAVPTTITTECSAATNAAGETTSVCSAKTIVSSASAGNTAPSAT-- 1231
QY 2330 VQSPQTRIRPST--PSQLSPGOQOQVTT--TSQPIPIQPTSLQIPQGOPOSQPV 2384
DB 1232 --PVTITAI-PTVITTESSVGTNSAGETTVGTATKSIPTTYITTL-IPESNAKKNETV 1287
QY 2385 QSSTQTLSSQOTLNOVSVSSPSRPOLOIQOPQOVIAVPOLOQOVYVLSQIQSVVAQIO 2444
DB 1288 ATAT-----NPISIKTTS-----QIATRASASVAPV 1315
QY 2445 AAGSGVPOQIKLOLPIQOQSAVQTHQIONV-VTVQAS 2483
DB 1316 TSPS-----LTGPELOSAGSNAVATVSPISSTIYGAA 1348

RESULT 9
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC 002817; Q14878;

FT	CONFLICT	1351	1351	H -> L (IN REF. 3).
FT <th>CONFLICT</th> <th>1412</th> <th>S (IN REF. 3).</th> <th>T -> S (IN REF. 3).</th>	CONFLICT	1412	S (IN REF. 3).	T -> S (IN REF. 3).
FT <th>CONFLICT</th> <th>1449</th> <th>L -> P (IN REF. 3).</th> <th>L -> P (IN REF. 3).</th>	CONFLICT	1449	L -> P (IN REF. 3).	L -> P (IN REF. 3).
FT <th>CONFLICT</th> <th>1504</th> <th>1504</th> <th>M -> T (IN REF. 3).</th>	CONFLICT	1504	1504	M -> T (IN REF. 3).
FT <th>CONFLICT</th> <th>4152</th> <th>4152</th> <th>G -> S (IN REF. 2).</th>	CONFLICT	4152	4152	G -> S (IN REF. 2).
SQ <th>SEQUENCE</th> <th>5179 AA;</th> <th>540295 MW;</th> <th>85CDV571IB945663 CRC64</th>	SEQUENCE	5179 AA;	540295 MW;	85CDV571IB945663 CRC64

Query Match	2.6%;	Score	391;	DB	1;	Length	5179;
Best Local Similarity	22.58;	Pred.	No. 1.6e-06;				
Matches 221; Conservative	91;	Mismatches	418;	Indels	254;	Gaps	33;

QY	1492	PSTKSDSDGQTPTPASQSCESNSVNOVEDMELETSVKVSSPLTSEESLMSDNDFIDE	1551
Db	3460	PTTPTTTTTTTTTVPT---PTPTGQTPTPTPTTTTTTYPPTPTPTGQTPTT-----	3508
QY	1552	NCGLPINKMENV-----NGESKRKTVLIEVMTST---VATESKTVLAKVGDQKQTV	1601
Db	3509	--TPTTTTTTVPPTPTPTGQTPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTTTTVPPT	3566
QY	1602	SSTENCASVTTTTTTATKSLSTPSTGSDVLISKEOSKIVTTTVDLSLTGCT--L	1659
Db	3567	PTPTGQTPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTTTTVPPTPTPTGQTPTPTPI	3626
QY	1660	VYSMTSVKSEYSTRDKVLMKFKSPKKTSGCALPSYKAFVKSSTKSIILVLPNDLKLTA	1719
Db	3627	TTTTTIVT-----PTPTPTGQTPTPTPTTTTTT-----	3654
QY	1720	RKGGIREVPYFNAAKPALDIMPSPSPPTFGITWRYLQTVKSLAGVSLMLRLMASLR	1779
Db	3655	-----VPTPTPTPTGQT-----QPTPTPTPT-----	3673
QY	1780	WDDMAKVPBGGSRTET--SETELTTTHPIIKRDVGPRGIFECIRKILCPIGP--E	1836
Db	3674	---TTTTTVPPTPTGQTPTPTPTTTTTTVPPTPTPTPTGQT-----PTPTPTPT	3720
QY	1837	TPKETPTPORKLRSSALRP-----KRPETPKQTG-----PVILETVAEELELMEI	1884
Db	3721	TTTTVPTPTPTGQTPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTTTTVP-----	3771
QY	1885	RAFAERVEKKAQVBOQAKRLEBOQPRVITSTTSPIS---STTSTISP-----AQK	1935
Db	3772	-----PTPTPTGQTPTPTPTTTTTTVPPTPTPTPTPTPTPTPTPTPTPT	3804
QY	1936	VMAVAPISGSVT-----TGTKMWLTTKGSPATYTPQONKNEHOFATWAWOGOSNGV	1988
Db	3805	PTTPTTTTTTVPPTPTPTGQTPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTTTT---	3861
QY	1989	VOVOQKVLGIIIPSSGTSGOQFTSFQOPRTATVIRPNTSGSGTSSNSGVITGPOIRPGM	2048
Db	3862	-----VPTPTPTGQTPTPTT--TPTTTTTTVPPTPTPTGQTPTPTPTPTT-----TTTTV	3908
QY	2049	TVIRTPLOOSTGKAKIRPVPVVOGAPQVQWTOILRGQPVSTANASAPNTVSTPOQKSL	2108
Db	3909	PTPTPTGQTPTPTPTTTTTTVPPTPTPTGQT-----PTPTPTTTTTTVPPTPTPTPTGQT	3963
QY	2109	TSATSTSNIOSSASOP--PRPOQGVKLMFAOLTOLGHHGNGQGLTVVLOGGQVGTGOLQ	2167
Db	3964	QPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTTTT-----VTPPTPTPTGQTPTTT	4015
QY	2168	LIPQGVVLPGGGOLMQAMNGTVQRFLETPLATTT-----ATTASTTT	2212
Db	4016	PTTTTTTVPPT-----TPTGT--QPTPTPTTTTTTVPPTPTPTGQTPTPTPTTTT	4066
QY	2213	TTVSTTAAGTGQRQSKSLPQMOVODKTLPPAOGSSVGPAPAKAQOTAPASRPQQTQRP	2272
Db	4067	TTTTVPTPTPTGQTPT-----TTPTTTTTVPPTPTPTGQTPTPTPTTTTTTT	4114
QY	2273	OSPAPQEVOTQOEVOQTQTVSSHPVSEAPQPIHAOSKPKOVAOSQPOQSVWQOSPVRAVS	2332
Db	4115	VTPPTPTPTGQTPTPTPTTTTTTVPPTPTPTGQT--PTTPTTTTTTVPPTPTPTGQT	4172
QY	2333	PSQTRIRPS--TPSOLSPGOOSOVQVTTTSPILPIQPHNTSLQIPSOQOPQ-----SOPQV	2384

Db 4173 PTTTPTTTTTTVPPTPTGTGTGTGPPPHNSTAPIALTLTNNPPPESTPQTSRSSSPLT 4223

QY 2385 QSSQTQLSSGQTLNQVSSPSRP 2408

Db 4233 ESTLLSLTPPAIEMTSTAPPSTP 4256

RESULT 10

AKA9_HUMAN	STANDARD:	PTT: 3911 AA.
AC	Q99996; Q9U004; Q9U0H3; Q9Y6Y2; O14869; O43355; O94895; Q9Y6B8;	
AD	16-OCt-2001 (Rel. 40); Created	
DT	16-OCt-2001 (Rel. 40, last sequence update)	
DT	16-OCt-2001 (Rel. 40, last sequence update)	
DE	A-kinase anchor protein 9 (protein kinase A anchoring protein 9)	
DE	(PKRA) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)	
DE	(Hyperion protein) (Yotiao protein) (centrosome- and golgi-localized PKN-associated protein) (CG-NAP).	
GN	AKAP9 OR AKAP450 OR AKAP350 OR KIA0803.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A. (ISOFORM 4).	
RC	TISSUE=Brain;	
RC	MEDLINE=98151389; PubMed=9482789;	
RA	Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.,	
RT	"Yotiao, a novel protein of neuromuscular junction and brain that	
RT	interacts with specific splice variants of NMDA receptor subunit	
RT	NRL."	
RL	J. Neurosci. 18:2017-2027(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.	
RC	MEDLINE=99219864; PubMed=10202149;	
RC	Witczak O., Skalhogg B.S., Keyser G., Bornens M., Tasken K.,	
RA	Janzen T., Oerstaak S.;	
RT	"Cloning and characterization of a cDNA encoding an A-kinase anchoring	
RT	protein located in the centrosome, AKAP450."	
RL	EMBO J. 18:1858-1868(1999).	
RN	[3]	
RP	SEQUENCE FROM N.A. (ISOFORM 3).	
RC	TISSUE=Brain;	
RC	MEDLINE=99287934; PubMed=10358086;	
RA	Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.,	
RT	"Characterization of a novel giant scaffolding protein, CG-NAP, that	
RT	anchors multiple signaling enzymes to centrosome and the golgi	
RT	apparatus."	
RL	J Biol. Chem. 274:17267-17274(1999).	
RN	[4]	
RP	SEQUENCE FROM N.A. (ISOFORM 1).	
RA	Kemmer W.A., Deiss S., Schwarz U.;	
RT	"Cloning of Hyperion."	
RL	Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.	
RN	[5]	
RP	SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).	
RC	TISSUE=Gastric parietal cell;	
RC	MEDLINE=99115654; PubMed=9915845;	
RA	Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,	
RA	Trotter K.W., Milgram S.L., Goldenting J.R.;	
RT	"AKAP350, a multiply spliced protein kinase A-anchoring protein	
RT	associated with centrosomes."	
RL	J. Biol. Chem. 274:3055-3066(1999).	
RN	[6]	
RP	SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).	
RC	TISSUE=Lymphoblast;	
RA	Hinds K., Sutterer C., Becker M., Hawkins M.;	
RL	Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.	
RN	[7]	
RP	SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).	
RC	TISSUE=Lung;	
RA	Milgram S.L., Goldenting J.R., Schmidt P.H.;	

```

    )
    anchor
  )
  ized
  i;

  g M.;
  tat
  t

  choring

  Ono Y.;
  that

```


DR PDB; 1MNE; 17-AUG-96.
DR PDB; 1VOM; 23-DEC-96.
DR PDB; 1LYK; 28-JAN-98.
DR DictyDb; DD01008; mhca.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head_1.
DR Pfam; PF00612; IQ_2.
DR Pfam; PF02736; Myosin_N_1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head_1.
DR SMART; SM00015; IQ_1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
FT DOMAIN 762 791 IQ.
FT NP_BIND 817 2116 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 638 660 ACTIN-BINDING.
FT DOMAIN 738 752 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678 ALKYLATION (SH-1).
FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MCK).
FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MCK).
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1EE56A1 CRC64;

Query Match 2.5%; Score 374.5; DB 1; Length 2116;
Best Local Similarity 19.1%; Pred. No. 2.3e-06;
Matches 356; Conservative 293; Mismatches 635; Indels 577; Gaps 81;

QY 31 DDDSDYBEEMDDDDSYCESSF--RSHSTYSSTPGCRKPRVRRSPLE----- 81
DB 530 DEQSVFNP-----ATDNTLITKLHSHFS--KKNKAYEERFSKTEGVTGTHYA 574
QY 82 -----EKDIPPLE-----FPKSEDLAVPNEHINVAIVEVLNFGTVLRLS 124
DB 575 GQVWELEIOMLEKKNKDLQDLCELFKSSDSDNYTKLFNPENISRAKKGANFTTV--- 630
QY 125 PFRFEDCALVSOEQCTLAEMHVVLLKAVLREDTSNTEG-----PADLKDS 174
DB 631 -----AAQYKQDLASLMTL-----ETTNPHFVCIIPNNKQDLPAKLEDK 670
QY 175 VNSTLYFDGTMPEVLR-----VYCESDKVHHVLPY--QEAED----- 212
DB 671 V-----VDOLRCNGVLEGIKITRKGFPPNRIITADYVKRYLLAPNVPRLAEDSQKATDA 725
QY 213 -----YRYGPVENKIKVLQFLVDQFLTNIAREELMSGV--IQYDHCRCYCHK 259
DB 726 VLKHLNIDPEGRFG-----ITKIFRAGQGLARIEBARQISIIIALQATGCGWLARK 780
QY 260 LGDLLCCEITSAVYHLECVKPLLEVEPEDEMQCVCAHKVPGVTDCAVLQCK-KPYIR 318
DB 781 -----VY-----KOAREHTVAARI-----IQQNLRAVI- 803
QY 319 HEPIGYRSRKRYWFL--NRRLIEDTENENKKIMYSTVKQVLAELIDCDKQWEA 375
DB 804 -----DFKSPWKKLSKAPPLKRNFEK----- 829
QY 376 ELCKILEEMREIHRHDITEDLTNKARGSNKSLAANEELIESIRAKKGDIDNVKSPE 435
DB 830 -----IKKERILFLKSNLIDSTTK-----DKLEKSLKDTESNVLDLQRL 872
QY 436 ETEKDKMETENDSKD---AEKNREEFDOLEKSDDKTPDDDEQKSEPEPTEVGKGN 492
DB 873 KAEKETIKAMDSKDALEAQKRELEIRVEDMESELEK-----K 911
QY 493 SVSNLQDNTNTAISETSSEGRSPVGCSETPDSSNMAEKYVASLPDVPPEPKTK 552
DB 912 LALENL--QNKRSVEEVRDLEE-----LOEQKLRNLTLE-KLKRYEELLEEMKRVND 964

QY 553 ESSNTSATTTSIOPNLNSNSSSELNSSQESAKAADDPENGRESHTPVSIQEIYGVDF 612
DB 965 GQSDTIRLEKIKIDELQ--KEVELTESPSESK-----DKVLEK--TRRILQSEL--DDL 1015
QY 613 T--SEKSTGELSESFGAGKAGSGSTRILTRLRNDSKLSQKSOQVAAAHAANKLFKEG 670
DB 1016 TVRLDSETKDKE-----LLRQKRLLEELKQVQVLAETAKILAQEA 1059
QY 671 KEVLVNSQGEISRLSK--KEYIMKGNINNYEPLSGEGKRYRYHNQYTNNSPALNHOH 728
DB 1060 ANKKL--QGEYTELNKEFNSEVTAARSV-----EKSRTLESQLVAVNNELEDEKKN 1109
QY 729 REDHDKRRHLAHKFCITPAGEFKWNG--SVHSGYL-----TISTRLITPOL 774
DB 1110 RDALEKKKALDALAEKMQQLESTGEEKSLYDLAKYQKQSDMEALRNQJSELSQTIAKL 1169
QY 775 ENNIPSSFHPHNMASHRANMIKAVQMCSPREPALALALIECAVYPVWLPPIREFLGHT 834
DB 1170 EK-----IKSP-----LEGV-----A 1181
QY 835 RLHMTSIEBEKEKVKKKKKOE--EETMQQA--TWKYTEPVKHQVKKQGEERYR 889
DB 1182 RLQGELEAEQDLANSVNEKQKKVELDEKSAQIAETAAKQALDKLKKLEQLELSEVO- 1240
QY 890 TGYGGMWIKTGHYRFPKLPQNTYN--YRKSLGTNNMDENMDESDKRCRSRSP 946
DB 1241 -----TQLEBAN-----NKNVNSDSTNKHLETSPNNKLLELAQKQKQKQALEK 1284
QY 947 KIKIIPSEKDEYKGDPAKAGADONE--MDISK--ITEKKDQVYKELLDSQPCKEEP 1002
DB 1285 RLGLESELKLVNEDLEBEKKQKESNEKKRYDLEKXVSELSQDIEEBAKKAAYTEAKN- 1342
QY 1003 MEVDDMKKTESHVQCQSSQVDVYVNSEGF--HLRSTYRK-----KTVYSKLQGLER-- 1053
DB 1343 -----KKSELDDELIRQYADAVYSSRKQSVQDLKTLQAKNEBELRNTAEABEQDLRAER 1395
QY 1054 --RIKQPTLE-----EKORLEKIKLEGIKGI-----GKTSNNSKNLSSES----- 1092
DB 1396 SKKAEFDLEBAVKNLEEFATAKVKAEKAMKKAETDYSRTKSELDAKNVSSQVQYQIKR 1455
QY 1093 -----FYITKAKGCOSDSKROQSPNANDQEDLIQCGSOSDSVLAWS----- 1138
DB 1456 LNEBELSELRYLEADERCSN-AIKARKTAESALESLEKIDIDANNAKAAEKRSLEY 1514
QY 1139 -----DPSHTTKK--LYPKDRLVDV-----SIRSPETK--CPKQ----- 1169
DB 1515 RVABLESLEDKSGTVVVERITRKDAEIDDLRLARLDRETSRKISDDEKNTKQFADLE 1574
QY 1170 -----NSIENDIEEKVSDIASRGQEPFKSKTGNDFFIDSK-----LA-- 1208
DB 1575 AKVEAQREVVATIDRLKKLESIDILISTQLDIETKSRIK-----LEKSKKLEQTLAER 1629
QY 1209 -----SADDLGITLCKANKRPLIOE--ESDTIYSSKSALHSS-----V 1244
DB 1630 RAAEEGSSKADE-----BIRQVQWEVDLEPRAQDLSEBAALANASAKIKTSIYAEVDEY 1683
QY 1245 PKSTNDR-----DATPLRAMDEF-----GKLGDSESNSTLENSSD--TVSIQDSSEED 1292
DB 1684 KEQLEDEIILAKDLVYKAKKRALVELEVRQDLBEEDSRSLEDSKRLTLTEVEDIKKKY 1743
QY 1293 MIVQNSNESISEQFRTREQDVEVLEPLKCELVSGESTGNCEDRLPYKGTENGAKKSPSQ 1352
DB 1744 DAEVQNTKLEDAKKKRLTIDVDPLK-----KQLEDEKKKLINSE 1782
QY 1353 KLEERPYNKCSQDQIKLNTIDKKNNEKRESEKKQGRSTQINGKDNKPKTYIKGECLEK 1412
DB 1783 RAKKRIESENEDPL-AKLDAEVKNRSRAEKDR-----KYEKDLKDTKYK--LNDAAARK 1834
QY 1413 ISRSRYVSGNVEPKVNNINKIIPENDIKSLJTVESAIRPINDQVIMEDFNENSSETKS 1472
DB 1835 --TQTEIGAARLEQIDELSLRSLKEQAKATQADKS--KTLLEGEIDNLRQIIEBGKIKM 1891

Oy		1473	HLLSSDA-EGYNRSLELTPSTKSDSTQTTTSPASCPDSNSVOMEDMETETSE--v	1528
Dd		1892	RLEKERRLLEGLELETRETFEAEDESKS-----EAOSKRIVELELEDARRLNQ	1940
Oy		1529	KKVTSPTSEESENLSDFIDENB-LPINKNEVENGSKRTVTTEYTMSTPATESK	1587
Dd		1941	KEIDAKELAEADKSNLQGEIYEAKGRLEEESTAFRNSDRKKRLAELDALTQAODAEC	2000
Oy		1588	TVIKYEKDKOTVVSVSTENCAC---STVTTTTTVTKSLSPSTGGSYDIISVEQSKT	1642
Dd		2001	AKNQIKENKRIETEELKEVKRKFGSESKTKTEFLVELVELETDYRAKKAEBDQQRLT	2060
Oy		1643	V 1643	
Dd		2061	V 2061	
RESULT 12				
ID	ATRX_MOUSE	STANDARD;	PRT; 2476 AA.	
AC	061687;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Transcriptional regulator ATRX (X-linked nuclear protein)			
DE	(Heterochromatin protein 2) (HP1 alpha-interacting protein) (HP1-BP38 protein).			
GN	ATRX OR XNP OR HP1BP2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NM	NCBI_Taxid:10090;			
NN	[1]			
NN	SEQUENCE FROM N.A.			
RP	MEDLINE=98213653; PubMed=9545503;			
RX	Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.:			
RA	"Comparison of the human and murine ATRX gene identifies highly			
RT	conserved, functionally important domains";			
RT	Mann. Genome 9:400-403(1998).			
RL	[2]			
NN	SEQUENCE OF 325-1176 FROM N.A.			
RP	MEDLINE=97133299; PubMed=8978696;			
RX	Le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,			
RA	Jeanmougin F., Losson R., Chambon P.:			
RT	"A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic			
RT	control of transcription by nuclear receptors.";			
RL	EMBO J. 15:6701-6715(1996).			
NN	[3]			
NN	SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC			
RP	HETEROCHROMATIN.			
RP	MEDLINE=20040663; PubMed=10570185;			
RX	Bickmore W.T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,			
RA	McGovern L., Pombo A., Turley H., Gatter K., Picketts D.J.,			
RA	Buckle V.J., Chapman L., Rhodes R., Higgs D.R.:			
RT	"Localization of a putative transcriptional regulator (ATRX) at			
RT	pericentromeric heterochromatin and the short arms of acrocentric			
RT	chromosomes.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).			
CC	-!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES			
CC	GENE EXPRESSION BY AFFECTING CHROMATIN.			
CC	-!- SUBUNIT: PROBABLY BINDS EH2. BINDS ANNEXXIN V IN A CALCIUM AND			
CC	PHOSPHATIDYLCHOLINE/PHOSPHATIDYLERINE-DEPENDENT MANNER (By			
CC	similarity).			
CC	-!- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC			
CC	HETEROCHROMATIN DURING INTERFERON AND MITOSIS. PROBABLY BY			
CC	INTERACTING WITH HP1.			
CC	-!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.			
CC	-!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL, AF026032; AAC08741.1;
DR EMBL, X99643; CAA67962.1;
DR MGD, MGI:103067; Xnp.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; Smp2_N.
DR InterPro: IPR001841; Znf_ring.
DR Pfam; PF00176; Smp2_N; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HelicC; 1.
DR SMART; SM00184; RING; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;
KW Zinc-finger.
KW ZN_FING 219
FT NP_BIND 1579 1586 ATP (POTENTIAL).
FT SITE 1704 1707 DEGH BOX.
FT DOMAIN 319 322 POLY-SER.
FT DOMAIN 735 738 POLY-SER.
FT DOMAIN 1001 1004 POLY-SER.
FT DOMAIN 1130 1135 POLY-GLU.
FT DOMAIN 1182 1185 POLY-SER.
FT DOMAIN 1238 1245 POLY-SER.
FT DOMAIN 1484 1487 POLY-ASP.
FT DOMAIN 1924 1931 POLY-GLU.
FT DOMAIN 2205 2208 POLY-SER.
FT DOMAIN 2245 2248 POLY-LYS.
FT DOMAIN 2403 2408 POLY-GLU.
SQ SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;

Query Match 2.5%; Score 373; DB 1; Length 2476;
Best Local Similarity 17.7%; Pred. No. 3,1e-06;
Matches 362; Conservative 300; Mismatches 658; Indels 726; Gaps 87;

13 AETDSEDEDEDEDEDEDD-----DSDYEEEMEDDDDDSYCTESSFRSHSTYSNP 65
26 SESESETCSSPLVNNQSTDKICGSLNSDM--WENKKEGASTSEKSRSSGSSRS--- 79
66 GRRKPRV-----HRRSPFILEEK-----DIPPLEPKSSEDLMPNEHIMNYI 108
80 -KRKRSIYTKVYESDDEKPTDENVNEKATENSENDITMQLPKGT--VIYQPEVLNE- 135
109 AIVELVRFNGVILRLSPRFEDFCALVYQEOCTMAEMHVVLLKAVLREEDTSVTTGCP 168
136 -----DKDDFKRPEF-----RSRSKKMKAD-----NLIRKGEDGLHGIYSC 170
169 ADLKQSVNSTLFLIDGMFWPEVLRYYCESDKRYHNVLRYQEAEDVRYQVEKKIKVLQFL 228
171 TAGGQOVN-----HFQDSTYRKH-----PSLKVLIDCKNC 199
229 VDOFLTTIAREELMSEVIOYDHGCRVCHKIGLDLCETCSAVVHLCV-----KPPLE 283
200 FKYYMSDDISRD--SDGM--DEGCRMCACBEGNLICDPCFNACCKKICILRLNGRKELS 253
284 EYVED--EMQCEVCAHKVPG-VTDC--VAE-----IQKNRPYIR----- 318
254 TIMDENNMOWCYICQPEPLDLVTACNSVFENLEQLLOQNKKIKVDESKTSKYVCDQTSK 313
319 -----HEPIGYDSRRKRYMVLNRLRIIEEDTEEN 348
314 FSPKSSSSNGEEKKLESCSGSVSYSHALSVPRE---MIKTTRLIETTSNMN 368
349 EKKIIVYSTKVLAELIQLDKDYWEALCK-----LLEEMREIIRHMIDTLEDTLTKA 402
369 SSYIKFLQAADNSGMTAM-----KLQCLSKSKSVLDDIK--AHLALEDL----- 414
403 RGSNKSFLAAANEEL--DESIIRAKKG--DIDNVYSPEETEKDKNETENDSKDAEKNREEF 458
415 -----NSEIQAADDVHKENETDLKSTDPKSKSTLKGKG-----KSYSTERKEEF 458

```



```

QY 459 -----EDQSLEKSDSDDKTPDDDPQSGKSEPT-----EVDGKNSVSA 496
D 459 LKLDARSSVKAIDGEEBRAHSTGEHKSGRGDSQOYEPTNTPEDLDMDIVSPSSVPE 518
QY 497 NLGDNNTNANSEETSPS-EGRSVGCLESTPDSS---NMAEK-----KVASEL 540
D 519 DIPSLSAMBYOSSADYQDNGSGTPELLESSVYKLVNSKDSRGNIKSVAKYAKVEL 578
QY 541 -----PODVPEEPNKTCESSNTSATTTSTIOPNLNSSSEL-----577
D 579 FVKLLTPVSLNSPIKGYVDCOEVSQEKNGR- KSSGVARSSEKCRPREISDHENNVITLLE 637
QY 578 NSSQSEAKAADDPENGERSHRPVSTQIETVGDFTSEKSTGELSESPGAKGASGSTR 637
D 638 DSDLRSPRYKTPPLRKQTSNPNAMSND-----EESNGTYKKR-----676
QY 638 IYRLRNPDSKLSQLKSOQVAAAHAANKLFKEGKREVLVNSQGEISLSTKKEVIMKNI 697
D 677 -QKASGPIRKDKRNSADCATDNPKPKHVPAKQPIGDNDSDEMLAVLKEASQMGHS 735
QY 698 N-----NYRKLQGECKYRYHHNYSTNSFALNKHQREDHDKRHLAHFCLTPAGEFK 751
D 736 SSSDPTDINEPQMHGK-----TGKDNNGK-----K 762
QY 752 WNGSVHGS-----KVLITSLRLITQLENNIPSSFPHPWASHRAMIKAVQMSKPRE 806
D 763 RKNSTSGSDPTKKGKSTETSIISKRRQVNSSESS---NYDSELENEITMSIGAKR 818
QY 807 FALALAILCAVFPVMLPIWREFLGHTRLHRMTSIREKEKVKYKKEKKQ-----857
D 819 -----SVP-----EKKEEDSSDEDEKGGKKVYDNGG 843
QY 858 -EEBETMOQATWVYTFPPVKHQVWKQGEYRYTVGYGMSWIKTHYRVPKLPFG-----912
D 844 HERAKTTQEGSSADDTGDT-----GRQGGSCSIAG-GSIEKVSQGVFERELCKPQVSSD 898
QY 913 -----NTNVN-----YRK-----SLEGT-----KNNMDENMD 934
D 899 GAEKPSVKEENVSPEDKRVSKTEKTKHLRSQRKKGSGSDGTDRFPKKEQSDSESE 958
QY 935 ESDKRKCSRSRKKIKIIPDSKDEVK-----GSDA-----AKGDQVQEMD 974
D 959 GEKQSRQRPQTKKAPDLGELTKREQWESSSDGTERLPEELETGIPSPKSIQKSDT 1018
QY 975 IS-----KITEKKDDVKELLDSDSDKPKCEPMVEYDDMKESHVNCQESSQVAVN 1027
D 1019 TAGGEKKGKKKKKCKCKEKKELSDSVKLPKGGSCDSSSEKTRNNVSLREKKRFSLLPA 1078
QY 1028 VSEGFHLRTSYKKRTKSSK---LDGLLRRIKQFTLEBKQRLKIKLEGIGIKIGTSTN 1084
D 1079 KSPKREPCSSSDTEKSLKGCQCCDS-TEKRPKRIDLRERR-----NSSSKRNTK 1126
QY 1085 SSKNLSESPVITAKREGQSDSMROEOSPANNNOQPEDLIQGCQSQSSSVLRMSDPSTT 1144
D 1127 EYKSHSSS-----SNAEGSSSDNKKQKKO-----RTAKKKTG 1159
QY 1145 NKLYPKDVLVDVSI RSPETKCPRONSLIENDIEEKVSDLSRQGEPTKSKTKGDNFTDD 1204
D 1160 N-----TEKKRNSLRAPPKRQVDITS-----SSSDIGDD 1191
QY 1205 SKLA-----SADIGILCKNNKP-----LIQESPTIYSSSKSAHSSVPKTNDRD---1222
D 1192 QNSAGEESDSE-----QIKIPVTENIYLPSTHTFCQSSGGEALSKSPALVAVDDDDND 1245
QY 1253 -ATLSLRAM--DPEGKIGCQSESNSTLENSDVTYSDSSEEMIVONSNESTSEQFT 1308
D 1246 PENRIAKKMLIEITKANSSDEGSSDDE-----PDGGGKRRTIGQSESPADGCEL 1297
QY 1309 REODVEVLEPLKCELVSESTGNCEDRLPYKGTANGKKPKQOQKL-----E 1355
D 1298 RREGOLAVNQ-----VNSSESDSDEE-----SKRPYRRRLRLRHKLTLTLDGSEGE 1341

```

```

QY 1356 EREPVNKCSDQIKLKNNTTDDKNNENRESEKGGQRTSTFOINGKDNKPKIYLKGECLKEISE 1415
D 1342 EKPT-----KPKHKEAKGNRR-----KVSSEDSSEDTDQESGVSEVSE 1382
QY 1416 SRVYSGVNEPKVNNINKIIPENDIKSLYKESAIRPPIINGDVIMEDPNERNSSETKSHLL 1475
D 1383 S---EDEQRRPRTSARKAELEENORSYKOKKKRRRRKQVED-----SSSENKSH--1428
QY 1476 SSSD-ABGNRYD-----SLETLTPRKESDSYQTTTPASCPESNS 1514
D 1429 SEEDKKEGDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1488
QY 1515 VNQVEDMELETSEVKRY-----TSSPTSEESNLSDNFDENGDLINKNENVNGES 1566
D 1489 RKRLAERERERKRLREVTEIEDASPTKCPITTK-----LVLDENE-----1529
QY 1567 KRRTVITEVTMTSTYATEKTYIKYKKGKQVYVSTENCASKSYTTTTTYTKLSTPS 1626
D 1530 -----TKEPLOVHRNMYIKLPHQVDGVOGEMDCCES-----VEKTKRSPGS 1573
QY 1627 -----TGGSVDIISVKEQSKTVTTVTDSLTGGLV-----TSMTVSKY-----1669
D 1574 GCLIAHMGIGKTLQVYSF-----LHTVLCDKLPSTALVYCPILNTALNMNNEFEKMO 1627
QY 1670 ---STRDKVLMKFSRPKKTRSGTALPSTRYKSTKSIYVLPMDLKLARKGCIIE 1726
D 1628 EGLDNMEKLEVSSELAIVKR-----FOERSYM-----LQRMQEDGVM 1665
QY 1727 VPFYFN 1732
D 1666 IGYEM 1671

RESULT 13
PHP_DROME STANDARD: PRT: 1589 AA.
AC P39769;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polyhomeotic (proximal) chromatin protein.
GN PH-P.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Braachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=imaginal disks;
RX MDLINE=92146957; PubMed=1346609;
RA Decanilis M., Cheng N.S., Pierre D., Brock H.W.;
RT "The polyhomeotic gene of drosophila encodes a chromatin protein that
RT shares polyene chromosome-binding sites with Polycomb.";
RL Genes Dev. 6:223-232(1992).
RN [2]
RP SEQUENCE OF 199-1584 FROM N.A.
RX MDLINE=92039031; PubMed=1937015;
RA Beatrick J., Daly M., Randsholt N.B., Brock H.W.;
RT "The complex genetic locus polyhomeotic in Drosophila melanogaster
RT potentially encodes two homologous zinc-finger proteins.";
RL Gene 105:185-195(1991).
CC - FUNCTION: BINDS TO POLYENE CHROMOSOMES. SEEMS TO INTERACT WITH
CC PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER
CC COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND
CC SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF
CC OTHER PAIR-RULE GENES SUCH AS EVE, FTY, AND H.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: SALIVARY GLANDS.
CC - SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC - SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X63672; CAA45211.1; -
 DR EMBL; M64750; -; NOT_ANNOTATED_CDS.
 DR PIR; S23632; S23632.
 DR FlyBase; FBgn0004861; ph-p.
 DR InterPro: IPR001660; SAM.
 DR Pfam: PF00536; SAM; 1.
 DR SMART; SM0054; SAM; 1.
 DR PROSITE; PS0105; SAM-DOMAIN; 1.
 KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
 FT ZN_FING 1365 1387 C4-TYPE.
 FT DOMAIN 1513 1577 SAM.
 FT DOMAIN 74 80 POLY-GLN.
 FT DOMAIN 411 450 GLN-RICH.
 FT DOMAIN 494 520 GLN-RICH.
 FT DOMAIN 619 650 GLN-RICH.
 FT DOMAIN 775 960 GLN-RICH.
 FT DOMAIN 1233 1290 SER/THR-RICH.
 FT CONFLICT 254 254 MISSING (IN REF. 2).
 FT CONFLICT 1415 1415 D -> A (IN REF. 2).
 FT SEQUENCE 1589 AA; 167297 MW; A6DF0CF9106E1891 CRC64;

Query Match Best Local Similarity 2.5%; Score 371; DB 1; Length 1589;
 Matches 255; Conservative 142; Mismatches 447; Indels 358; Gaps 50;

QY 1758 LQTVKSLAGVSLMLRLMLASLRMDMAKVPG-----GGSTRTESEETITTEIKR 1811
 DB 93 LFTLAQAGITPEDEK-----DVASPRHPGLAQOQATSGCPATGSSVPTPS--HR 142
 QY 1812 RDVGPVIREYCIKIKICPIGVPEPKETPTPOKGLRSSALPKRPETPKOTGVILE 1871
 DB 143 HGTPPTGR-----QHTTP-----STPNRSPAPSTPNFNCSTI- 175
 QY 1872 TWAVEELEMEIRAFEREREKQAQAVEQAKRLEQKPYIAITSTPT----- 1923
 DB 176 -----ARHTSLTLKKAQNPQO-----VAAITVPLQISPOLQ 209
 QY 1924 ----STTTSTISPAQKVVAVPISGVTGTMVLTTKGVSPATVFOONKRFHOTFATWK 1980
 DB 210 QFYASNPYAIQVAKQEPPTHTSSG--GTELKHAINT--HEVQQLDLOQLSPANG 261
 QY 1981 QGQSNQGVVOQKVLGIIPSSSTGSOQFTTSPOPTAVITRP-----NTSGSGGTTN 2035
 DB 262 GGAASAGAGGAS--PANSQSQOQ--QHSFAISTSPMLAATGAGGDMTQ 311
 QY 2036 SQ-----VITGPQIRPGMTVIRPLOQSTLGAIIITPVNPGARQOV 2079
 DB 312 GRIVVQLMOPSTSLYRPMIVSGLMLHFG-GLGQOPLOVITAGPFP-----GQNGPQL 363
 QY 2080 MT-----QIRQO-----PVSTAVSAPMTVSTSPQKSLTSATST 2114
 DB 364 TTTTQNAKKMIGGAGPAGSNVATCIPTNHNSPQVLFSPMNVISPOQONLQS---- 419
 QY 2115 SNIQASASPPRQOGQVKITMAQLTQLDGHGNGGLVIVIOGOG-QTTGOL-QLIPQG 2172
 DB 420 ----MAAAADQOQLTQOQOQFNQOQOQOQLTQO--QOQDLTFAALKVGVDAQGLAKQVVK 473
 QY 2173 VIVLPGPGQOILMOAMPNGT-----VORLFTPLATATATASTTTTPTTAGTGE 2224
 DB 474 VTTT-----SSAVALATPGSTGTQTQOQVQVQOQOQTTQTQOQVQVSQSTLPVGVGG 529
 QY 2225 QRSKSLSPQMVQVHODKTLPPAQSSSVGPAKAQ--PQTAPSARPP-----QTQP-- 2272
 DB 530 Q-----SVQTAQLNLNAGQAQOQMLPWFLOMAAQLQPGPQMIIIRNQPDG 574

QY 2273 -----QSPAQPEVQTPF--VO-----TOTTVSSHVSEAPPHNAQSSKPOVAQSQPS 2320
 DB 575 TQGFMEIQOQAPQTLQIQOQNIQOCNVTPTRKRTQDLALAPKQOQOQOQVGTNTQOQ 634
 QY 2221 NVQGSFVRVQSPQSTNT-----RPSRP-----SOLSPQSQSQVOTTSQPI----- 2362
 DB 635 QQLAVATAQLQOQOQOQLTAAALQRPAGPVMPHNGTQVRPSSVSTQTAQMSILKAKMRN 694
 QY 2363 ---PIQPTSLQIPSGQPOPSOPV-----QSSSTQ----- 2389
 DB 695 KQGPVRPALATFKTEIGQVAGQNKVNGHLITVQOQOQATNLQOVVNAAGKMYNSTGT 754
 QY 2390 --TSSQQTILNYSVSSPSRPOLDIQOPROVAVPD--LQOVYVLSQIQSOVVAQIQ 2445
 DB 755 PTLQNGQTLHAATAAGVDKQOQOQLFLQKQQLIQOQOQMLQOQITAAITOMQOQAAVQAOQ 814
 QY 2446 QQSGVPQKILQPLQIQOSSAVQTHQIQNVVTVQAASVQEQQLQRYQQLRQOQKQKQO 2505
 DB 815 QOQ---QVSGOQOVNAQOQOQVAAQOQ-----QAVQAQOQOQREQOQVAAQAOHQ 864
 QY 2506 IEINVN-----TPSKLLIREVETIQOVNKHNAVIEHLKQKS-----MTPARRE 2551
 DB 865 ALNATQQLIQVANPQFTSHQ--QOQOQQLHNLQIQOQQAQVQAQVQAQAOQOQ 922
 QY 2552 NQRMIVQNYMKYILDKIDKEKQAARKKRESVEQKSK----- 2592
 DB 923 QQREOQNIILQIQIVQOQSGATSOQTQOQOQHQSGLQLSVFPSSVSTTPAGIATSSA 982
 QY 2593 -QNAFKLSALFKHKEQ-----LRAELIKRALLDLDQIQLEVEDELKRLDKIK 2639
 DB 983 LQALSLASGAIPOYAKPGCTGSSSPSSVVTITNOSTPLVTSSVIAQOQOQSAQVH 1042
 QY 2640 KENDLMO--LAQATAVAAPCP-----VTPVLAPAP-----PSPPPPGVQHTGL 2685
 DB 1043 QHQQLISATLTAGGQOQOPQPSLFTPTNPLAMTSMNATVGLSLAPVTVSYSTAV 1102
 QY 2686 LSTP---TLVASQ---KRRKEEKSSSKKKMKMISTSEK--KDTKLYCIK 2734
 DB 1103 TSSPGQLVLLSTASSGGGSIPTPTKETPSKGPATLVPDIGSKPTVSGKDT-----CT 1157
 QY 2735 TP 2736
 DB 1158 TP 1159

RESULT 14
 HRX_HUMAN
 ID Q03164; Q14845; Q16364; Q13743; Q13744; Q9UMA3;
 AC 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN MLL OR HRX OR ALLI OR TRX1 OR HTRX.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=9304667; PubMed=1423624;
 RX Tkachuk D.C., Kohler S., Cleary M.L.;
 RA "Involvement of a homolog of Drosophila trithorax by 11q23
 RT chromosomal translocations in acute leukemias";
 RL Cell 71:691-700(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96290553; PubMed=8703835;
 RA Nilsson I., Loechner K., Sieglar G., Grell J., Beck J.D., Fey G.H.,
 RA Marschalek R.;
 RT "Exon/intron structure of the human ALL-1 (MLL) gene involved in
 RL Br. J. Haematol. 93:966-972(1996)."

RN [3]
 RP SEQUENCE OF 1-1909 FROM N.A.
 RX MEDLINE=93390935; PubMed=8378076;
 RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
 RA Koderia Y., Nakawara S., Ariyoshi Y., Takahashi T., Ueda R.;
 RT "Two distinct portions of Utrig19/ENL at 19p13 are involved in t(11;19)
 RT leukemia.";
 RL Oncogene 8:2617-2625(1993).
 RN [4]
 RP SEQUENCE OF 1317-2328 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93265134; PubMed=1303259;
 RA Djabali M., Sellieri L., Parry P., Bower M., Young B.D., Evans G.A.;
 RT "A trithorax-like gene is interrupted by chromosome 11q23
 RT translocations in acute leukaemias.";
 RL Nat. Genet. 2:113-118(1992).
 RN [5]
 RP SEQUENCE OF 1251-1538 FROM N.A.
 RX MEDLINE=94215165; PubMed=8162575;
 RA Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaan O.,
 RA Saito H., Croce C.M., Canaan E.;
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 RT involved in acute leukemia.";
 RL Cancer Res. 54:2326-2330(1994).
 RN [6]
 RP SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RX MEDLINE=95322025; PubMed=7598802;
 RA Mongkolkeha D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
 RA Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 RT trz zinc-finger domain, and alternative splicing.";
 RL DNA Cell Biol. 14:475-483(1995).
 RN [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE=95315013; PubMed=7794749;
 RA Marschalek R., Grell U., Lochner K., Nilsson I., Siegler G.,
 RA Zwackromer I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 RT transcripts in the acute lymphoblastic SEM cell line with chromosomal
 RT translocation t(4;11).";
 RL Br. J. Haematol. 90:308-320(1995).
 RN [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE=94020842; PubMed=841518;
 RA Forster A., Rabbits T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 RT application to isolation of MLL fusion CDNAs from acute leukaemia
 RT translocations.";
 RL Oncogene 8:3157-3160(1993).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAST.
 RX MEDLINE=20183971; PubMed=10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAST translocation early during
 RT chemotherapy with DNA topoisomerase II inhibitors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 RN [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
 RX MEDLINE=20115194; PubMed=10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;
 RT "Novel SH3 protein encoded by the AF3P21 gene is fused to the mixed
 RT lineage leukemia protein in a therapy-related leukemia with
 RT t(3;11)(p21;q23).";
 RL Blood 95:1066-1068(2000).
 RN [11]
 RP FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -1- DISEASE INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL.
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
 CC THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES

CC MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 CC MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;
 CC T(11;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)
 CC that involves MLL and ELN; t(11;19)(q23;p23) that involves MLL
 CC and GAST; and t(3;11)(p21;q23) that involves MLL and AF3p21.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.
 CC DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MLL.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L04284; AA58669.1; -;
 DR EMBL; Z69744; CAA93625.1; JOINED.
 DR EMBL; Z69745; CAA93625.1; JOINED.
 DR EMBL; Z69746; CAA93625.1; JOINED.
 DR EMBL; Z69747; CAA93625.1; JOINED.
 DR EMBL; Z69748; CAA93625.1; JOINED.
 DR EMBL; Z69749; CAA93625.1; JOINED.
 DR EMBL; Z69750; CAA93625.1; JOINED.
 DR EMBL; Z69751; CAA93625.1; JOINED.
 DR EMBL; Z69752; CAA93625.1; JOINED.
 DR EMBL; Z69753; CAA93625.1; JOINED.
 DR EMBL; Z69754; CAA93625.1; JOINED.
 DR EMBL; Z69755; CAA93625.1; JOINED.
 DR EMBL; Z69756; CAA93625.1; JOINED.
 DR EMBL; Z69757; CAA93625.1; JOINED.
 DR EMBL; Z69758; CAA93625.1; JOINED.
 DR EMBL; Z69759; CAA93625.1; JOINED.
 DR EMBL; Z69760; CAA93625.1; JOINED.
 DR EMBL; Z69761; CAA93625.1; JOINED.
 DR EMBL; Z69762; CAA93625.1; JOINED.
 DR EMBL; Z69763; CAA93625.1; JOINED.
 DR EMBL; Z69764; CAA93625.1; JOINED.
 DR EMBL; Z69765; CAA93625.1; JOINED.
 DR EMBL; Z69766; CAA93625.1; JOINED.
 DR EMBL; Z69767; CAA93625.1; JOINED.
 DR EMBL; Z69768; CAA93625.1; JOINED.
 DR EMBL; Z69769; CAA93625.1; JOINED.
 DR EMBL; Z69770; CAA93625.1; JOINED.
 DR EMBL; Z69772; CAA93625.1; JOINED.
 DR EMBL; Z69773; CAA93625.1; JOINED.
 DR EMBL; Z69774; CAA93625.1; JOINED.
 DR EMBL; Z69775; CAA93625.1; JOINED.
 DR EMBL; Z69776; CAA93625.1; JOINED.
 DR EMBL; Z69777; CAA93625.1; JOINED.
 DR EMBL; Z69778; CAA93625.1; JOINED.
 DR EMBL; Z69779; CAA93625.1; JOINED.
 DR EMBL; Z69780; CAA93625.1; JOINED.
 DR EMBL; D14540; BAA03407.1; -;
 DR EMBL; L01986; AAA92511.1; -;
 DR EMBL; U04737; AAA18644.1; -;
 DR EMBL; S78570; AAB34770.1; -;
 DR EMBL; X83604; CAA58584.1; -;
 DR EMBL; S66432; AAB28545.1; -;
 DR EMBL; AF231998; AAG26532.2; ALT_TERM.
 DR TRANSFAC; T02337; -;
 DR Genew; HGNC; 7132; MLL.
 DR MIM; 159555; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003889; FYrich_C.
 DR InterPro; IPR003888; FYrich_N.
 DR InterPro; IPR003616; POSTSET.

DR InterPro: IPR001214; SET.
DR InterPro: IPR002857; ZnF_CXXC.
DR InterPro: IPR001965; ZnF_PHD.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR Pfam: PF02008; ZF-CXXC; 1.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50280; SET; 1.
DR PROSITE: PS01359; ZF_PHD_1; 3.
DR PROSITE: PS50016; ZF_PHD_2; 3.
KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;
Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
KW Alternative splicing; Repeat.
FT DOMAIN 17 102 ALA/GLY/SER-RICH.
FT DNA_BIND 169 180 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 217 227 A.T HOOK (BY SIMILARITY).
FT DNA_BIND 301 309 A.T HOOK (BY SIMILARITY).
FT ZN_FING 1147 1194 CXXC-TYPE.
FT ZN_FING 1431 1482 PHD-TYPE 1.
FT ZN_FING 1479 1533 PHD-TYPE 2.
FT ZN_FING 1566 1627 PHD-TYPE 3.
FT DOMAIN 1703 1748 BROMODOMAIN (DIVERGENT).

Query Match 2.5%; Score 367; DB 1; Length 3969;
Best local similarity 17.6%; Pred. No. 8.7e-06;
Matches 620; Conservative 405; Mismatches 1211; Indels 1278; Gaps 153;

QY 5 EEEEDGDAEETODSDDEDEDEDDDDSDYEEEMEDDDDDSYCTEES----- 54
Db 856 EELSKRDADKSVKSKRREREREK--RESREKKKSGEIQSSALYPYGRVSK 913
QY 55 -----FRSHSTYSTPGRRKPRVH----- 73
Db 914 EKVGEDVATSSAKKATGKKSSSHSDGTDITSVLTGDTAVKTKLLIKKGNLEKTN 973
QY 74 ---RPSPIIEEDIPPLEPKKSEDLMPNHEIMVIAIYEVLNFGVLRISPEREED 130
Db 974 LDGPTAPSLTEKREKTLCLSTPSSS-----TVKHSTSSIGSM----- 1010
QY 131 FCAALVSOECCTLAEMHYVL-----KAVLEEDTSNTTF- 166
Db 1011 -----AQAQDKLPMDKRYASLLKKAQOLCKTEKSKLKQTDQPKAQGGESDSETSVR 1064
QY 167 GP-----ADLKSYNSTLYFDGMTWPE---VLRYCESDKREYHHVLP 206
Db 1065 GPRIKHVCRRAVALGRKRAVFPDDMT---LSALPWBEREKILSMGMDDS--SIAG 1118
QY 207 YQEAEDYPRGPEVNAIKVLOFLVDQFLTNIAEELMSEGVIOYDHCVRCHLGLDLC 266
Db 1119 SEDAE--PLAPPIKPIK-----PVTRNKKAPQEPVKKGRSRRCQ-----C 1158
QY 267 EPGSAVYHLECVKPLPEVPEDEMOCEVCAHNVPGVTDVAIQNKPIRIHEPIGYDR 326
Db 1159 PGC-----QVPED---CGVC-----TNCL-----DKP----- 1177
QY 327 SRKKYFLNRRLIIEEDTENENKIMYYSTKVQLAELIJDLDKDYWEALCKILEEMRE 386
Db 1178 ---KFGGRN-----IKQCKCKMKKCNLOMPS 1202
QY 387 EIHHRMDITEDLTNKAAGSKSFLAANEILISIRAKKGIDINVKSPFEETKDKMETEN 446
Db 1203 KAV-----LQKQAKAVK-----KEKSKTSEKKDSKSSVYKKNVDS 1240
QY 447 DSKDAKNEEFEDSDDKTPDDPEQCKSEPTVGVGKGSVSNLGDNTTAT 506
Db 1241 SQKPTPSARD-----PAPKSSSEPPPKRPFVEKSEB-----GN-VSAP-GPESQAT 1287

QY 507 --SEETSPSEGRSPVGLSETPDSSNNAEKKVASSELPODVEEPPNKTCESSNTSATTTSI 564
Db 1288 TPASRKSSKQVOPALVIPPOPTGPBR-----EVKPTTPEPKK----- 1330
QY 565 QPNIENSSSELNSSQSEAKAADP-----ENGRESHTVYSIOELIVGDTSEKST 618
Db 1331 QP-----PPESGPESQKKAAPRSIPVKQPKREKRPVKNQEN----- 1373
QY 619 GELSESPGAKGASGRITRLRNPSKLSQLKSOQVAANAHEANKLPKGEKVLVNS 678
Db 1374 -----AGTINILSTLSNGNS-----KQIPLAGVIRIRVDFEDCEAEVWE 1416
QY 679 QGEISRLS---TKKEVIMKGINNYFKLGQEKYRVYHNOYSTNSPALNKHOREDHDK 734
Db 1417 MGGIGITVSPITPRVVC-----FLASSGHVEEVQVOCCEP----- 1455
QY 735 RRIHAKFCITPAGERKKNSSVHGSKVLTITLRLITTOLENNIPSSFPPNASHRANW 794
Db 1456 ---HKFCL-----ENERPLEDQLENNCCRC-- 1479
QY 795 IKAVQMSKRRERFALALALEC-----AVKPVVMLPIWREFLGHTPLR 838
Db 1480 -KFOHVG--ROHQATKQLECKMCKNSYHPECLGPNYPKPKKKKWTI-----CTKCVR 1532
QY 839 MTSIEREKEKVKKKKKQEEETMQATWVYTFYKQVWK--QKGE----- 885
Db 1533 CKSGSGTTPK-----GMDAQW-SHDFSLCHCAALFKAGNCPLCDKCYDD 1578
QY 886 ---EYRVTYGWS-WIS-----KTHVYFVPLDGN---TWNVRK----- 920
Db 1579 DYESKMQQCKGCDRWVHSCENLSDMEYELISLDESVAVYTCVNCTERPAEMRLALEK 1638
QY 921 ---SLEGTNNMDENDESDKRCSPKIKIEPSE-----KDEVKSD-----AA 965
Db 1639 ELQISLQVLTALNLSSTTSHLLRYROAKPPLDLPNETESIPSRSPGCDPPLVETVS 1698
QY 966 KGADQNMEDISKITEKKQD-----VKELDSDSDKPKCEPEVDDMK 1010
Db 1699 KODQDQPLDEGVKRRKMDQGNYSVLEFSDDIYKIIQAALNSDGGQ-----EIKK 1749
QY 1011 TESHVQCQESSQYDVYVNSGFIHRTS-YKKTKSKSLDGLERRIKQPTL----- 1060
Db 1750 ANSMVKSFFIRQWE--RVFVFWISVKSREWEPPNKVSSNSGMLPNALVPLSLDINYAQOM 1807
QY 1061 -EKKQRLKIKLGGI-----KGIGKTSNNSKNLSESVIIRKAKGCGSDSMDROSP 1113
Db 1808 REENSHTEQPLMKKIIIPAKPKGPEPDSPTPLHPTPTPL-----STRSRKDSR 1859
QY 1114 NAN-----NDQPEDLIQ-----CSQSDSSVLRMSD----- 1139
Db 1860 ELNPPGIEDNRQALCLTYGDDSANAGRLLYIGQENWTHVACALMSAVFEDDGSLLK 1919
QY 1140 -----BSHTTNKLYPKDRLVDYSINSPEPKCRKONS 1171
Db 1920 NVHMAVIRGKOLRCERCQKPGATVGCITSTSNYHMCSTRANCVFLDKKVVQCORHRD 1979
QY 1172 I-----ENDIE-----EKVSDLAS--RGQETSKTKTGDNFFIDD-----SKLAS 1209
Db 1980 LKGEVAVPENGFEVFRVAVDFEGISILRRKFLNGLRENIHMIGSWTIDCLILDLSD 2039
QY 1210 ADD-----IG-----TLICK--NKKPLIQ-----EESDITVSSS 1236
Db 2040 CEDKLEPIGQCSRNVWSTTDARKRCVYICKIYECRPVYEPRLINSVEHDEKRTIAHSP 2099
QY 1237 KSAIHSSVPKSTN-----DR-----DATPLSRAMF-----EKKLGC 1268
Db 2100 TSPTSSSKESQNTAELIISPSPDRPHSQTSGSCYUHVISKVPRIRTPSYSPFORSPOC 2159
QY 1269 D---SENSSTLENSDVIYSDSSSEDMIVQNSNEGISQORTREQDVEYLER--LACELY 1324
Db 2160 RPLPSAGSPPTTHETIVGDP-----LSSGLRSIG---SRHSTSSLSPPRSKLRIR 2210
QY 1325 SGESTNCEDRLPVKCTEANGKRPQOK--KLEERPVNKCSDQIKLKNITDCKNN----- 1377


```

Db 2211 SPARTGTWYSNNVSVSTGTATDLESsAKVVDHVLGLPLNSSTLQONTSTSNLQRTV 2270
QY 1378 ---ENRSEKKGQRTSTF-QINCKDNPK--IYLKGECLKEISERVAVSGNVEPKVN--- 1429
Db 2271 VTGNNKNSHLDGSSSSSEKOSASDLVSKSSSLKGEKTKVLS-----SKSEGAHNAY 2325
QY 1430 --INKIPE-NDIKSLTVKESAIRPTNGDYMEDFERNRS-SETKSHLLSSDAENYR 1485
Db 2226 PGIRKLAPQVHNHTSRRLNYSKISGFAPSSV--SFSSKELSPHLHLNRQRDRROHT 2383
QY 1486 DSELTLPSTKESDS-TOTTPSASCPSNSVNO-----VEDMEIETSE 1527
Db 2384 DSTQSSNSPDEDETVTLKLISGMNSRSSIINEHMGSSSRDRROKGRKSCKEFEKHS 2443
QY 1528 VKKVTSSPITSEESNLSNDEIDE-----NGLPIKNE 1560
Db 2444 KSFLPEQVYTTGEGCNLKPEPMDEVLPETWGPFCNNVSDKIGDKGLSPGVKAPPM 2503
QY 1561 NVNGESK-----RKVTTEVTMTSTVATESKTVIKVEKDKOTVVSSTENCASKSTVTT 1614
Db 2504 QVESAKELQAPRKRYKVTLLPLKMNESQSKNALK-----DSSPASPLOIESTS 2554
QY 1615 TTTTIVTLSTPSTGSGVDIISVKQSKTVVTTTDTSLTTGGTLVTSMTVSKETSTRDK 1674
Db 2555 PTEPISASENPGDG-----PVAOPSPN--NTSCODSOSNNYOML-----PVQDR 2596
QY 1675 VKLKFRPRKTRSGTALPSYRKFTVKTSKISFVLPNDLKL--ARKGIREVPYF-- 1730
Db 2597 -NMLPQPKRQEDGSKFRKRPKRSARASRMF-----GLTPLYGVSIGEDIPFYSS 2650
QY 1731 -----NYNAKPAIDIMYPSPRPFTGITWRYRLQTVKSLAGVSLML 1771
Db 2651 STGKKRKRSAEGQVDADLSTDEDDLYYNNFTK-----TVISSGEE-- 2695
QY 1772 RLMLASLRMDMAKVP-----GGGSTRTERSETEITTEILIKRDVGPICIREYCIRK 1827
Db 2696 RLASHNLFREEQCDELKISOLDGVDGTE--SDTSVAT----- 2733
QY 1828 IICIGVETPKETPTPOKRLGRSSA--LRPKREPTEKOTGPVLIETWAAEELMEIR 1885
Db 2734 -----TRKSOIKRKNCKENGTEMLKIDREPDAKEKHVTKSSVGAKNEFKMCH 2784
QY 1886 AFAERVEREKAQAEEQAKRRLKEQOKPVIATSTTSSTTSTISPAOKVMAPIGCV 1945
Db 2785 SVSR-----VKTOGQDSLFAQLSSL-----ESSRVHSTPSDKMLD-- 2822
QY 1946 TTGKRMVLTKVGS-----PATVTFQONKMFHOTFATWVQOGSNS-----G 1987
Db 2823 TYNTELLKSDSDNNNSDDCGNILPSDIMDFVLKNTPSMOALGESPESSSELNLGEGLG 2882
QY 1988 VVOYQOAVLGIIPSTGTSOOTFTSFQPRATVIRPNTSGSGGTTNSGVITGPQIRPG 2047
Db 2883 LDSRKEDMGF-----EVEFQQLPTEPV-----DSSVSSISAEQFELPLELPSD 2930
QY 2048 MTVIRTPLOOSTLKAIIRPVMVOPCAPOQVMTQIIRGQPVSTAASAPNTVSTPGOKS 2107
Db 2931 LSVLTJ-----RSPTV-----PSGNPSRL-----AVISDSEKR 2959
QY 2108 LTSATSNISQSSASOP-----PRPO-----QGV-----KITMAQLIQLTQGHG 2147
Db 2960 VT--ITKSVASSFSDALLSPGVDPPEGHMPPDHFIOGHMDADHISPPCGSVEGCHG 3017
QY 2148 GNOGLIVVIOGQGTQLOLIPQGVTV-----LPQGO----- 2181
Db 3018 NNQDLT-----RNSSTPGLQ--VPVSPVPIQONKYVNSNDSPEPSSISNAAVOTTPHL 3071
QY 2182 -----QIMO-----AAMPNGTVQRELFPTPLATATTASTTTTIVSTTAAGTEQ 2225
Db 3072 KPATEKLIVANNQNNQPLVYQLTLPNGVTKIQLTSSSVSPSVETMTVSLVGPBG--GLT 3130
QY 2226 RQSLTSQOMOVHODKTLPPAOSSSVGPak-----AOPQTAQSAAROPOTOP----- 2272

```

```

Db 3131 LTTGLNLSPLTSQ--SLFSPASKGLPMSHHQHLSPAPTAQSSFPNNISNPPSGLLIGV 3188
QY 2273 QSPAQPEV-----QOPPEVOTORTVSSHVS--EAOP--THAOSKPOVAAQOSOPQSNVOG 2325
Db 3189 QPPDPOLIVSESSQKRDLDSTYAT--PSSGLKRPISRLOTRKNKLALASSPSNLT--- 3243
QY 2326 SPVRVOS-----PSQTRIRSPSTQSLSPG-----OOSOVOTTTSQPI 2363
Db 3244 APSDVSNMMLINFTPSQL---PNNPSLLDLGSLNTHSHRTVPNIIKRSKSSIMYFEPAP 3300
QY 2364 IQPH-----TSIQIPSOQPOSOPOVOSTOTLSSGOTLNOVSVSPSPQLOIQP- 2415
Db 3301 LLPQSVGTAATAAGTSTISQDTSHLTSVGSVGLASSSVLNVMSQOTTTPTSSASVPG 3360
QY 2416 -----OPOVIAVPOLOQOVVLSQIOSOVAAQIOQOSGPOQIKIQLPIQIOQSSA--- 2467
Db 3361 HVTILNRLILGTPIGS-----ISNLIKASQSLGIQDO-----PALPPSSGMFP 3407
QY 2468 -VQTHQIONVTVQASVQDQRLQVQQLRDQOKKQOQIEINVPYPSKLLIKVEIIOKQ 2526
Db 3408 QLGTSQTPSTAAITPAASSICVL-----PSTQTTGITPAAPS----- 3443
QY 2527 VVMKHNAVIEHLKOKKSMPTAEREENORMIYCNVMYIIDKIDKEKQAKKRRRESV 2586
Db 3444 -----GEADEHYQLOHVNOL----- 3459
QY 2587 EOKRSKONATKLSALFLFKHKEQLRAELIKRRALLDKDQIEVQELKRDLIKKEKDLAQ 2646
Db 3460 -----ASKTGT-----HSSQ-----RLDLSASGQVSNFQYDAPRSMG 3494
QY 2647 LAQATAVAAPCPVTPVLPAPRAPPSPPPPGVQHTGLST--PVLVASQKKRKEEK 2704
Db 3495 LEQNKALS-----SAVASPTSPGSGSPSSSQRSASPSVGPPTPKPKRTKRFQLPLD 3548
QY 2705 DSSSKSKKKKKMISTSKET-----KKDKIKVICICTPYDESKFYICDRCQNMWHGRVGI 2760
Db 3549 KGNCKRKHVSHLRTSSSEAHPIPDQETTSLTSGTGPAAEA----- 3589
QY 2761 LQSEAEILIDEVYCPQOCOSTEDAMTVL-----TPLEKYEGLKRVLSRLQAHKMAPF 2813
Db 3590 -QDPTASVESSQKQEGQAPQAVALEVQYQONPABEQ----- 3527
QY 2814 LEPVDPADPDYGVIKEPMDLATMEERVQRRYKLTFFVADTKIFDNCRIYNPSDSP 2873
Db 3528 -ESAEPKTEVEEESNFSSPLMLMWLOQDQKRR---ESTTEKKPKKGLVFE-----ISSDDG 3678
QY 2874 FYQCAEVLSEFFVQKLGFKASRSHNNKLOSTAS 2907
Db 3679 FOICAESIED-----AMWSLTDKVOEARS 3702

```

RESULT 15

ID	NCRL_MOUSE	STANDARD;	PRT;	2453 AA.
AC	060974;	060812;		
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, last sequence update)			
DT	16-OCT-2001 (Rel. 40, last annotation update)			
DE	Nuclear receptor co-repressor 1 (N-COR1) (N-COR) (Retinoid X receptor interacting protein 13) (RIP13).			
GN	NCOR1 OR RXRIP13.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10909;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).			
RC	TISSUE=Plutitary;			
RX	MEDLINE=96008539; Pubmed=7566114;			
RA	Hoerlein A.J., Naeaeer A.M., Heinzl T., Torchia J., Gloss B.,			
RA	Kurokawa R., Ryan A., Kamei Y., Soederstrom M., Glass C.K.,			
RA	Rosenfeld M.G.,			

"Ligand-independent repression by the thyroid hormone receptor

RT mediated by a nuclear receptor co-repressor.";
 RL Nature 377:397-404(1995).
 RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Liver;
 RX MEDLINE=95280959; PubMed=7760852;
 RA Seol W., Choi H.S., Moore D.D.;
 RT "Isolation of proteins that interact specifically with the retinoid X
 receptor: two novel orphan receptors.";
 RL Mol. Endocrinol. 9:72-85(1995).
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
 ABSENCE OF LIGAND.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBICUITOUS.
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR
 INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 DOMAINS (ID1 AND ID2).
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND
 SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES
 FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 SPECIFICITY.
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CORN BOXES.
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U35312; AAB17125.1; -;
 DR EMBL: U22016; AAC52168.1; -;
 DR TRANSFAC: T04688; -;
 DR MGD: MGI:1349717; Ncor1.
 DR InterPro: IPR001005; MYB_DNA_binding.
 DR Pfam: PF00249; myb_DNA-binding; 2.
 DR SMART: SMO0395; SANT; 2.
 DR PROSITE: PS00090; MYB.3; 1.
 KW Nuclear protein; transcription regulation; DNA-binding; Repressor;
 KM Coiled coil; Alternative splicing;
 FT DOMAIN 174 216 COILED COIL (POTENTIAL).
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT DOMAIN 299 328 COILED COIL (POTENTIAL).
 FT DNA_BIND 437 482 SANT-A (POTENTIAL).
 FT DNA_BIND 619 669 MYB.
 FT DOMAIN 501 550 COILED COIL (POTENTIAL).
 FT DOMAIN 606 616 PRO-RICH.
 FT DOMAIN 2073 2077 CORN BOX OF ID1.
 FT DOMAIN 2277 2281 CORN BOX OF ID2.
 FT DOMAIN 58 64 POLY-GLN.
 FT DOMAIN 593 602 POLY-ALA.
 FT DOMAIN 1044 1047 POLY-PRO.
 FT DOMAIN 1718 1718 POLY-ALA.
 FT DOMAIN 1968 1979 POLY-SER.
 FT VARBLIC 2333 2371 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1952 1952 I -> T (IN REF. 2).
 FT CONFLICT 2090 2090 A -> P (IN REF. 2).
 SQ SEQUENCE: 2453 AA; 270640 MW; 52208BA0382F7E6A CRC64;

Query Match 2.4%; Score 365; DB 1; Length 2453;
 Best Local Similarity 18.0%; Pred. No. 5.9e-06;
 Matches 438; Conservative 296; Mismatches 782; Indels 912; Gaps 108;
 QY 298 HAYGVYDCVAEIQKKKPIIRH---EPIDGRSRKKYFNLRLRIEED----- 343
 DB 378 HEISEIIDISGEENNEKQROLVPIPMFDAQRRVFINNNGMEPMKYKRODM 437
 QY 344 ---TENENE-----KIIWYSTKQALALIDCLDKDWEALCLILEMREIHR 390
 DB 438 NWTDEKELEFKDFIQHPNENGLASYLERKSVPCVLYT---LYKKNEYKALVR 493
 QY 391 HMDITDLTNKAGSNKSFIAANEELISIRAKKGDIDNVKSPETEDKNETENDSD 450
 DB 494 NY-----GKRGRNQQLARPSQEE-----KVEEKEDKAEKTEKKEE 530
 QY 451 AEKNREFEDQLEKSDDKTPDDPEQKSEPT---VGDKGNSVANLGDNTNATS 507
 DB 531 EKKDEKEDD---KEDSKETTKEDRTETATAEPEEREQVTPRGKRTANSQGGKGRVTR 587
 QY 508 EETSPSEGRSPVGCLETPDDSSMAKKVASLPPQVPEEPKNTCESSNTSATTSIQN 567
 DB 588 SMTSEAAANAAATAEEP-----PPPLPPPE----- 615
 QY 568 LENSNSSELNQSQSESAKAADDPENGERESHTPVSIQEIYVGFSEKSTGELSEPGA 627
 DB 616 -----PISTEVEYSRWIEE-----EMEVA 635
 QY 628 GKAGASSTRITRLRNPDSKLSQQAQVAAAHAANKLFKSGEVLVYVNSGEISRLST 687
 DB 636 KKG-----LVEHGR-----NMAAIAMKGT 655
 QY 688 KKEVIMKGINNYFKLGQSGKYVYVHNOYSTNSFALNKQHHRHDHDKRHLAKKFLTPA 747
 DB 656 KSEAQCKNFPNT-----KRH----- 672
 QY 748 GEEKNMGVHSGKVLITSLRLITLQLENNIPSPFHPMASHRAMIKAVOCSKREF 807
 DB 673 -----NLDNLQO---HKQKASR-----KPRE- 691
 QY 808 ALALALECAVRYVNLPIYREFLGHTRLHRMTSIREKEKYKKKEQEBEETMQOAT 867
 DB 692 --ERDVSQ-----ESVAST-----VSAQEDDIEASNEENPESEGENSS 732
 QY 868 WVKYTPPVKHQVWKQGE---YRVYGYGWSWISKTVH-----YRFPRKLPQNT 914
 DB 733 DTE-SAPSPSPVEAKSSEDSSENAASRGNTPEVALEATTDPAPCASPSSAVP---TT 787
 QY 915 NVNYRKSLEG-----TKNMDENMDESDKRKCS--RSPKRIKIEPDSKDEYKGS 963
 DB 788 KPRERSEVEAQYTDASAFELAPMDVDHCEGAGESSVLDPAFPAKTADSVDPMEQVPEMT 847
 QY 964 AAKG-ADQNEMDI---SKITTEKKDQV-----KELDSQDKPKCEE----- 1001
 DB 848 ASKGEDDAKERDLESTSEKTEARDEDDVVAEQIERPEQSDSSATCSADEGVDEPER 907
 QY 1002 ---PMEVDDDKTE-----SH-----VNCQESSQYDVVNS 1029
 DB 908 QRVFPMDAKPSLTTPPGSLISSPIKIPNLLDPQLQHRAAVIPPMVSCYPCN-IPIGTV 966
 QY 1030 EGFHLRTSYKKTKSSKGLGLERRIKQFTLEERQYKLEKIGKIGKIGKSTNSKN- 1088
 DB 967 SGVAL---YQRIKAMHESALLE-----ERQROQDYDELCR--SSSPCSTKSPMR 1014
 QY 1089 ---LSESP--VITKAKESQSDSMRQESP-----NANNQPEDLIGCQSGSDSS 1133
 DB 1015 EMEVLPAPAHQVITNLPREVRLPTTRPPPLIPSSKTVAASEPSTIMGSGISQGP 1074
 QY 1134 VLKMSPTHTTKLYPKDVLVDVSIKSPETKCPKQNSIENDIEEVSLDASRGQPTPS 1193
 DB 1075 GTYLS--SH--NQAYPO-----EAPKPSVSGISLGL-----PRODESTKA 1110


```

QY 1194 KTKNDFEIDSKLASADIDGTLICKNKRPIL---OEESDTIVSSKSLALHSSVPKSTND 1250
Db 1111 -----APLTIYKOEFS-----PRQNS 1128
QY 1251 RDATPLSRAMD---FEGKLCDSSESNSTLJENSSDPTVSIODSSEEDMTVQNSNESISBQFR 1307
Db 1129 QPEBLVRAQHEGVVGRGTAGAVQGSITRGTPAKSIS----- 1165
QY 1308 TREDDEVELEPKCELVSCESTGNCEDRLPYKG--TEANGKKPSQOKKLEERPYNKCSQ 1365
Db 1166 -----VETISSLRGSIQGPRA-----LPQAGIPTALVKGPVSRNMPLEESSEPEKREE 1214
QY 1366 IKLKNTP--DKKN-----ENRESEKKGORT--STFOINGKDNPKI---YLKCECLKEI 1413
Db 1215 AASGCHVIEBKSGHLLSYDNIKARREGTSPRTAHMSLKRSEAVEGSIKQMSMRBS 1274
QY 1414 SESHVSGNVEPKVNNINKIIPENDIKSLVYESAIRPFIINGDIVM-----EDFNE-- 1464
Db 1275 PVSAPBLEGLI---CRALPRGSPHSDLKERTV-----LSGSIHQGTPRATASPFEDGL 1323
QY 1465 -----RNSSETSHLSSDAGCNKRDILETLPSTKESSTOTTPPSASCPESSNVQ 1517
Db 1324 KYPROIKRESPIRA--FEGAITKGKPYDGITTI--KEM-----GRSIEH 1364
QY 1518 VEDMEIETSEYKVV-----TSSPITSEESNLSDNDFIDENGLPIKNKNENVGESKRTVI 1572
Db 1365 IPRODILTOESKRTPEVQOSTRPII---EGSIS-----QCTPI-KPDNNSGQSAIKHNV 1414
QY 1573 TEVTMTSTVATESKTVIK-----VEKDKO-----TVVSTENCACKTV-- 1612
Db 1415 KSLITGPBKLPRGMLIEIVPENIKYVERGKYEDVAKAGEPVARHNTSVVSSGSPVLRSTLHE 1474
QY 1613 -----TTTTVTYKLSPTSTGGSV-----DIISVKEOS--KTVVTTITVD 1650
Db 1475 APKAOBSGLYDSDSARTPVSYOYTTISRGSPMANRTSDVSSKSASHEKSTLTPTORE 1534
QY 1651 SLTTTG-----GLIVT-----SMTVSKYSTRDVKVL--MKFSRPKRTSGTAL- 1692
Db 1535 SIPAKSPVGVDPVIVSHSPFDPHRRSSAAGEVYRSHL.PTHLDPAMPFHRALDPAAYLLO 1594
QY 1693 -----PS-YKFEVTKSKKKSIFVLPNDLKKIARGGIREVPYFNNAKP--ALD 1739
Db 1595 ROLSPPTGYPYOYLAYAMENRTOLL--NDYITSQOMQVNL--PDVTRGLSPREOPLG 1649
QY 1740 IMPYSPPTFGITWRYRLQTVKSLAGVSLMLRLMLASLRWDDMAAKVP-----GGG 1792
Db 1650 L-PYATRGITDLT-----NMPTIIVPHAGT 1676
QY 1793 STRETSETETTEIIRKRDVPGYGRFEYCIRKIICPIGVPEPKETPTPORKGLRSS 1852
Db 1677 STPMDRITTYIPGTQVT-----FPPRP-----YMAA 1702
QY 1853 ALRPRRPTPOTGPVILIEVVAEELELWEIRAPAEVEKEKAQAVBOAKKRLBOOKP 1912
Db 1703 SLSPEHP-----THLA-----AASAER-EREREKERERERERER- 1740
QY 1913 TVIATSTSTSTSTISPAQKVMAPISGSVTTGTAKMLLTKVGSFATVTFQONKNFH 1972
Db 1741 -----EBERIAAPADLYLRGSE-----QGRPGSHGYV----- 1770
QY 1973 QTFATWVQOGSNGCVVOYQKVLGIIIPS-----STGTSQOTFTSFOPRTATVITRPNTSG 2028
Db 1771 -----RSPSPSVRTOETIILQORPSVFOGINTGS--VITPLDP-TAOLRIMPLPSG 1817
QY 2029 SGGTNSGOVLTGPQIRGPMVITRPTLQOSTLGKAIITPVWQPGAPQVMTQIIRGQ 2088
Db 1818 -----GPSISOGLPASR-----YNTADALAL----- 1840
QY 2089 VSTAVSAPNTVSTPGOKSLTSATSTSNIOSSASQPPRPOGQVYKLTMAQLTQLTQGHGG 2148
Db 1841 VDAASAPQOMDVSKTKESKHEAARLEENILRSRAVVSBOQOLEQKNLEVE----- 1890
QY 2149 NOGLIVVIOGOGQTTGOLLOLIPQGVTVLPFGQQLMOAAMNGTVQRFLFTPLATTATTA 2208

```

```

Db 1891 KRSVQCVCTSSALPSSKAQ--PHASVYSEAGKD--KGPPKSRYEELRTRGKTTITTA 1946
QY 2209 STTTTIVSTIATACTGQORQ-----SKLSPQMCHQDKTLRPA-----OSSVGPAAQOP 2258
Db 1947 NFIDVITROIADKDAEREGSQSSDSSSLSHREYETASDAEVIISPASSPAPPOEKFO 2006
QY 2259 TAOP-----SARPOPTQPSQAPPEVOTQ-----PEVOTQ 2289
Db 2007 AYQPDWAKNQAQNESTROYEGPLHHYRQOESPPQOQPLPLPSSQSGMGQVPRTHRL 2066
QY 2290 TVVSHVPSAOPTHAOSKPOVAAOSQPOSNVQOSPVYRVOSPSQTRIRPSTPQSLSG 2349
Db 2067 ITLADHICQIITQDFARNQVPSQASTSTEQT-----SP--SALSSTPVRTKTSRYSP 2118
QY 2350 QOSQOVQTTQSPLP---IQPHTSLQIPSOQOPQOSQOPVQSSQTTLSSGQTLNVQSVSSPS 2406
Db 2119 SOSQ---TVLHPPRGVSPENLVDSKRSRPEKSP-----RSHIFS 2158
QY 2407 RPQLQIQOQPOQVIAVAPOLQOQOVVLSQ 2434
Db 2159 EPEEPIPPQGP--AVHEKQDSMLLSQ 2184

```

Search completed: November 20, 2002, 16:26:32
 Job time : 117.509 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:21 ; Search time 48.0411 Seconds
(without alignments)
5817.155 Million cell updates/sec

Title: US-09-698-295-1
Perfect score: 14971
Sequence: 1 MVSEEEEDGDAEETQDSE.....KIKGFKASRSHNNKIQSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4082	27.3	810	2 G01252	small GTP binding
2	854.5	5.7	1711	2 T21432	microtubule-assoc
3	546.5	3.7	5327	2 T13564	hypothetical prote
4	514.5	3.3	2447	2 T16870	hypothetical prote
5	490	3.3	5170	2 T15348	hypothetical prote
6	463.5	3.1	3507	2 T34513	hypothetical prote
7	462.5	3.1	3759	2 A35085	trithorax protein
8	458.5	3.1	3488	2 T34418	hypothetical prote
9	446	3.0	2897	2 B48666	cell proliferation
10	440	2.9	2271	2 F90073	hypothetical prote
11	439.5	2.9	3256	2 A48666	cell proliferation
12	439.5	2.9	3924	2 S37431	ankyrin 2, neurona
13	439	2.9	2938	2 T30249	cell proliferation
14	419.5	2.8	2232	2 T34434	hypothetical prote
15	417	2.8	2441	2 D71623	erythrocyte membra
16	407.5	2.7	3828	2 T13857	trithorax protein
17	406	2.7	4377	2 A55575	ankyrin 3, long sp
18	405.5	2.7	452	2 T21435	hypothetical prote
19	403	2.7	5105	2 T32650	hypothetical prote
20	402.5	2.7	510	2 T21430	hypothetical prote
21	401.5	2.7	405	2 T21433	hypothetical prote
22	401.5	2.7	1367	1 S48478	glucan 1,4-alpha-9
23	397	2.7	2722	2 T20532	hypothetical prote
24	397	2.7	6642	2 T29757	protein UNC-89 - C
25	393	2.6	1630	2 A53577	ascites stanolgyco
26	393	2.6	2481	2 D90011	fmrb protein flmpo
27	392	2.6	1871	2 D96796	probable heat shock
28	392	2.6	2738	2 E88320	protein F07A11.6 l
29	391	2.6	1939	2 T18372	repeat organellar

30	391	2.6	3147	2 T18674	hypothetical prote
31	383	2.6	2346	2 T13829	trp homolog - fruit
32	378.5	2.5	1229	2 T25697	hypothetical prote
33	377.5	2.5	2215	2 T16871	hypothetical prote
34	375.5	2.5	3259	1 A56539	giantin - human
35	375	2.5	2526	2 T20531	hypothetical prote
36	374.5	2.5	2116	2 A26655	myosin heavy chain
37	372.5	2.5	1589	2 T13606	hypothetical prote
38	372	2.5	1526	2 A45605	mature-parasite-in
39	368.5	2.5	3225	2 T52300	giantin - human
40	367.5	2.5	1829	2 T24583	hypothetical prote
41	366.5	2.4	2187	2 T30826	nascent polypeptid
42	366.5	2.4	4550	2 T18440	hypothetical prote
43	365	2.4	2453	2 S60254	nuclear receptor c
44	362.5	2.4	2843	1 RBH0AP	adenomatous polylo
45	361.5	2.4	2541	2 T29340	hypothetical prote

ALIGNMENTS

RESULT 1									
G01252									
Small GTP binding protein SEC4 homolog - human									
C:Species: Homo sapiens (man)									
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 29-Jan-1999									
C:Accession: G01252									
R:Bowser, R. P.									
submitted to the EMBL Data Library, January 1994									
A:Reference number: G06441									
A:Accession: G01252									
A>Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 1-810 									
A:Cross-references: EMBL:005237; NID:q1276427; PID:q451848									
C:Genetics:									
A:Gene: FAC1									
Query Match									
Best local similarity 98.7%; Pred. No. 3.8e-144;									
Matches 777; Conservative 4; Mismatches 2; Indels 4; Gaps 1;									
QY	1	MVSEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDASCTESSFSRSHST	60						
DB	1	MVSEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDASCTESSFSRSHST	60						
QY	61	YSTPGRKRPRVHRPSPILKEKDIPPLEPKSSDLPMPNEHIMVIAIYELRNFTY	120						
DB	61	YSTPGRKRPRVHRPSPILKEKDIPPLEPKSSDLPMPNEHIMVIAIYELRNFTY	120						
QY	121	LRSPRFEDFCALVSQECCTMAEMHYLLKAVIREEDTSMTTGPADLKQSVSTLY	180						
DB	121	LRSPRFEDFCALVSQECCTMAEMHYLLKAVIREEDTSMTTGPADLKQSVSTLY	180						
QY	181	FIDGMTPVELRYVCESDKEYNHVLRYQEAEDYPYGVENKIKVLOFLYDQFLTTIARE	240						
DB	181	FIDGMTPVELRYVCESDKEYNHVLRYQEAEDYPYGVENKIKVLOFLYDQFLTTIARE	240						
QY	241	ELMSEVIGYTDHCHRCVCHKLGDLCCETCSAVYHLCVAPRPLEVEPEDEKQCVCAHY	300						
DB	241	ELMSEVIGYTDHCHRCVCHKLGDLCCETCSAVYHLCVAPRPLEVEPEDEKQCVCAHY	300						
QY	301	PGVTDVAETQKPKYIRHEPIGYDRSRKRYWFLNRLLIEETDENNEKKIYYSTKYQ	360						
DB	301	PGVTDVAETQKPKYIRHEPIGYDRSRKRYWFLNRLLIEETDENNEKKIYYSTKYQ	360						
QY	361	LAELIDLDKDYAEALCKLIEEMREEIRHMDITEDLTNKANGSKSFLAANEELIS	420						
DB	361	LAELIDLDKDYAEALCKLIEEMREEIRHMDITEDLTNKANGSKSFLAANEELIS	420						
QY	421	IRAKKDDIDNVKSPETEEKDKNTENDSKDAEKNRFEFQDSLEKSDSKTPDDPEQK	480						
DB	421	IRAKKDDIDNVKSPETEEKDKNTENDSKDAEKNRFEFQDSLEKSDSKTPDDPEQK	480						


```
QY 1382 SEKKQRTSTPOINGKDNKPKIYLKGECLKEISEKSVNNGVNEPKVNNINKIIPENDIKS 1441
Db 1002 ----- 1001
QY 1442 LTVKESAIRPFINGVIMEDFERNNSSETKSHLSSDAEGNYRDLFTLPSTKESDSTQ 1501
Db 1002 ----- 1001
QY 1502 TTTPASCPESNSVQVEDMEIETSEVKKVTSSPITSEESNLNDPIDENGLPIKNEN 1561
Db 1002 -----SG 1003
QY 1562 VNGEKKRTVITEVTMTSTVATESKTYIVKKGKQIVVSTENCASKVTITTTTITVK 1621
Db 1004 VGEERK----- 1009
QY 1622 LSTPTGSGVDIISVKEGSKVTVTTTVDLSLTTGGTLVTSMTVSKEYSTRDKVKLMKS 1681
Db 1010 ----- 1009
QY 1682 REKTRSGTALPSYRKRYKSTKKSIFVLPNDLKLARKGIREV--PYFVNAKPAID 1739
Db 1010 -----AMPIEIQTFSFKRGKSGIFVLQKKILRQIMGGCQGYVMGFGAGIKSNLL 1062
QY 1740 IMPYSPRPFEGITWRYLQTVKSLAGVSLMLRLMASLRMDDMAKVPDGGGSTRFETS 1799
Db 1063 IMPYAPRPTLDCWKMQTFLNARSLHVAVALQKTIWSSIKFNEFD--DDTHPRDRVVI 1119
QY 1800 ET--ETTTEILKRVDVGYGIRFEICIKICPI-GVETKEKETPTPRKG-----LR 1850
Db 1120 DTPSHDERRRILRHKEMPYGYQERYEMEIEIIPLYDEEEDDESLSNRKSGSSEFSHR 1179
QY 1851 SSALPRKREP-----PKOT-----GVIIETVAEELELME----- 1883
Db 1188 SSARKKRRQRIEFLSLKGNTPKSKNAFRSLDNRRAIRREWDVGLKVFELKDW 1239
QY 1884 --TRAFAEREKEKQAV-----EQAKRLEQOKPTVIATSTTSPTSTSTISPAQ 1934
Db 1240 KMRLEAEATARKKLEATRKAKAKEDERRRLOOQOQSVARIPV-PHMS--LIPSE 1294
QY 1935 KMWVAPISSV-----TTGKMWLTTRKVSAPATVTPQN-----KNFQOTATWKKQO 1983
Db 1295 RNNV--PYLGSQOQRRPNGERGFLEKYNNSSSVSPQAHAYASTPRPGYQPOPNILRQAG 1353
QY 1984 SNSGVVQVQKVLGIIIPSSSTGTSOQFTSFOPRIATVITRPMTSGSGGTSNSQVITGPQ 2043
Db 1354 YNQ-----LPRKPTTSPFNFS-RP-VATIPITPQLRAAGADG--VIRAVM 1396
QY 2044 IRPG---MTVIRTPLOOSTLGAIIIRPVWVQPGAPQGVMTQIIRGQ----- 2087
Db 1397 MTPGNKSTVNTSTPYPQ-----ALNRQOYLOQOQOQCAVRRLTNGYHFMQDGMGGGR 1451
QY 2088 -----PSTVASANVTYVSTPGQKSLTSAT-----STSIQSS 2120
Db 1452 NPSVQHQLPQNRALQRFGESESTEMRVRTEALIPNDGDGQPVVIRPYDPTSNFDAQ 1511
QY 2121 ASQPRPQOGVQKLMMAQTLTOGHGNGQGLTVVIOGOGTTGOLQ-----IPQCVTVL 2176
Db 1512 RAOQOHPOGRPVYSTPAQIKRTTQPGVKH--NVILMAKSDGTQKVVLPKPGCFPTVI- 1568
QY 2177 PGPQOQLQMAAMPNGSTVQRELTPLATTATSTTTTSTTAAGTGEORQSKL----- 2230
Db 1569 -STGGRVIVYRQPTAVQOQLYV-----ATPGTRVVRIPNANGAGAPROQDHQVMRRVQ 1621
QY 2231 --SPQMQVHQDK--FLPRAQ-----SSSYGPKAKPQRTQPFARQPOQVQPOS 2274
Db 1622 ASGPRAEMEMDQGTTPPQOQRYVYLQGGNSGTPVNVNPKVSSRGPRGLTMQVQOQO 1681
QY 2275 PAQPEVOTQPEVOT---OTTVSSHVPEEAO 2301
Db 1682 QHNPAHYDMPDQATGFAVSTTTTQVPEDEQO 1711
```

```
RESULT 3
113564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
A:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.7%; Score 546.5; DB 2; Length 5327;
Best Local Similarity 18.3%; Pred. No. 4e-12;
Matches 550; Conservative 479; Mismatches 1154; Indels 825; Gaps 121;

QY 2 VSEEEEDGDAEETQDS-----EDDEE-----DEMEEDDDSDYPEEMEDD 44
Db 2191 VQDEAKQKRSREKRESVIPPRAKDKSPKVELQPVSMTEITRELDADQPMKPSQASRR 2250
QY 45 DD-ASYCTESSFSRSH-----STYSSTPGR-----RKPRV-----HNRSPFILE 82
Db 2251 ESIAESIKASPRDEKSPASKASRPGSAESIKYDLDPQIKDKDKSTHSRRSLED 2310
QY 83 KDIPLEPFKS--SEDLMVNEHNMVIAIYEVLRNGVLRLSPFREFD-----PCALIV 136
Db 2311 KSAVTSE--KSVSRPLVASDH--EAAVALEDKAS--STSP--KQSRGFEAFETV 2359
QY 137 SQ--EECTLMAEHVYLLKAVLR-----EEDTSNTTFGPADLKDVSNTLYFTD 183
Db 2360 SSTIEATMEFKIEVEKSSIALSIQGGSGGKLQTDSSPVVAEEDPFHAVAASVTTP 2419
QY 184 GMTWP-EVLRYVCESDKEYHNVLPQEAEDYPGP-----VENKIKVLQ 226
Db 2420 TLTKPAELAQIGAKTVS--SPIDELALRTPSAPRHISRADSPACASEIASQDKSPQ 2475
QY 227 FLYDQFLTNIRAEELMSGVLYQDDHCVRKHLGDLLOCCEFSANVYHLECKVP----- 280
Db 2476 VLKSSRPAMVAESK-----DDAQLKSSVEDLR--SPVASTEISRPASAGET 2521
QY 281 ---PLEEVEPEDEMOCEVCYAHKVPQVTDCAVETQKKPKYIRHEPIGYDRSRKRYFLNRR 337
Db 2522 ASSPIEAKRDAEFP--QAER--AVLPRLTELKGLP----- 2555
QY 338 LIIEDTENENKIKIYVSTKVLQALIDCLDKDWEALCKILEMREIHRHMIDTED 397
Db 2556 -----TLSPVDAHASVQPAELSKV-----DIEKT 2581
QY 398 LNKAKGSKSPFLAANERELISIRAKKGDIDNVKSPETEEDKNTENDSKDAEKNRRE 457
Db 2582 ASSPIDEAKSLIGSPAEE-----RPESPASADAAVSVERKSADASPPSV 2628
QY 458 FEDQSLKSDSD-----KTPDDPPOGK--SEPTVEGDKGNSVANT--GDNTTNAV 506
Db 2629 VESTKADSTKGDISPESVLEGRKDVKSKESSRP-----SVASITGDSTKDV 2681
QY 507 -----SEETSPSEGRSPVQCLSETPDGSNNAKKVASELPQDVPEEPNKTCESN 556
Db 2682 RPASVYESVKDEHDKAESRESIAKVESVIDEAGKSDSKSS--QDSQKDEKSTLAKRE 2738
QY 557 TSNTTTSIOPNLNENSSSELNSSQSESAAKADDPENGRESHITPVSIOE-----EIVVD 611
Db 2739 ASKRESV---ESSKDAEKSESRESVIAAGEPV--PRESKPLDSKDTSRPGSVES 2792
```



```

Db      4579 EEADADVDRSSKESRSISATIMMTSTYKPSEDM--EPISK--LVEEHE-HVELAQ 4633
Qy      2634 RDLKIKKEDLQAQATVAAPCPVPVVL-----PAPPAAPS 2673
Db      4634 EYTSRSTKTTLLQSSQSTSTTSSTKSGASRVESITTLQMDQTSQSGDPADKRTPTT 4693
Qy      2674 PPPPGVQ 2681
Db      4694 APVSPGV 4701

RESULT 4
T16870
hypothetical protein T13H2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C:Accession: T16870
R:Wu, X.
Submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid T13H2.
A:Reference number: 218593
A:Accession: T16870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2447 <WGX>
A:Cross-references: EMBL:U39653; NID:q1049397; PID:q1049400; PIDN:AB52494.1; GSPDB:GNOC
A:Experimental source: strain Bristol N2; clone T13H2
C:Genetics:
A:Gene: CESP:T13H2.3
A:Map position: X
A:Initons: 18/3; 135/3; 176/3; 496/1; 693/1; 816/1; 1053/1; 1131/3; 1181/3; 1308/3; 1645
C:Superfamily: RING finger homology
F:158-207/Domain: RING finger homology <RRN>

Query Match      3.4%; Score 514.5; DB 2; Length 2447;
Best Local Similarity 18.2%; Pred. No. 2,3e-11;
Matches 514; Conservative 423; Mismatches 1037; Indels 857; Gaps 118;

Qy      9 EDGDAEETQDSDEDEDEEDDDSDYPEEMEDDDDDASYCTESS----- 54
Db      2 DDSPPSTSKSARDAKEMNEMTSDSSSDSEVSASASESRPSEKKKVTITRVIPVP 61
Qy      55 -----FRSHSYSTPGRRKPRVHRPSPILKEKDIPLPEFKSSDLMVNEHIMNYI 108
Db      62 PTRDGHVNLLESGNESEETKSLYQRAKGIPIKPKPEIKLPTTSEQYDLEEVLM-- 119
Qy      109 AIYEVLNRNGYVLRSPRFEDPCALVSEQCTLMAEKNHVLLAKVLEEDTSNTTGP 168
Db      120 -----PAMEG-----RELTLNA----- 132
Qy      169 ADLKDVNSTLYFIDGMTPVELRVYCESDKREYHNHLYQEAEDYRGVFNKIKVLOFL 228
Db      133 ---YDAVRKKYVNLPG-----KSYCEAD-----LQKV 156
Qy      229 VDQFLTTNARELMSEGVYQYD--CRVGHKLDLLCETCSAVYHLECYKPPLEEV 286
Db      157 IGSF--SCDVCQCLIGSITTKCKGHRFCDCILVAFMRSGNTCPTCRQMLGSKRELQCP 215
Qy      287 E-DEMOCCEVCNKHVPGYVDVVAEIQKN-----KPIRIEPIGYDSRRKRYWLNKR 337
Db      216 RRDQLIYQVESRSIVG--RMAENNEHEKDYVFGRRKGIETG---GSDWNKRGGIDPNK 270
Qy      338 LTIEDTENEKKI-WYSTKVLQALIEDCLDKDYWEALCKILEEMREETIRHMDITE 396
Db      271 LKAPRPLKSAGRRKIRMFHES-----DED---GSVAKVMESKGA-----PKE 310
Qy      397 DLTNARSGNSKFLAANEELIESTRAKGDID---NVKSPETEKKNETENSKDAEK 453
Db      311 DDTNLENDKEGTVAAEKEVLE---EGEMDFPEIKSDEQCDLDEEESMDSDF 365
Qy      454 NNEEFEDSL-----EKDSDDKTPDDDPQGSSEETVEGDGNSVSAWL 498

```

```

Db      366 EISDNEDVSKPSCSTSKTNTNRSRDSSESDNDSRDELQKKRKK-----KKNVPKTD 419
Qy      499 GDNNTNATSEETSPSEGRSPVGLSETPDSSNMAEKKVASLEPOVPEEPNKTCSSNTS 558
Db      420 GSDVSNESFEDDASGE-----VATKL---IKESKKKPCGPKKK 456
Qy      559 ATTTSIQPNLENSNSSSELNSQSESAAKAD-----DPNGERESHTPVSIOE 606
Db      457 FAPELIEGDIPLPSEEDLSLSDDERDNDADPYAFVQKEFNDRP---RRDH----- 506
Qy      607 EIVGDTSEKSTGELSESGAGKAGSSTRITIRLNPNSKLSQLKSQVAAAHAHANKL 666
Db      507 -----PEKD-----KLYNFEMIDM-----NHQVDRK 528
Qy      667 FKEGKVLVNSOGELISRLSTKKEVIMKGINNYFLLQGEKRYVYHNQYSTNPFALNH 726
Db      529 FEKDGELHIYISDSSNEHSEDAE-----DRESSIDS----- 560
Qy      727 QHRREDHDKR--RHIAHKFCLTPAGEFKMNGSVHGSKVLTITLRLITOLENNIPSSFH 784
Db      561 ---EHEKEISKFLSHRQPLNPNTSVDDCOY-----ITVAKD----- 595
Qy      785 PNWASHRAMWIKAVQCSKPREFALALILECAVRYVWMLPIWKEFLGTRLHRMTSIR 844
Db      596 -----VKQSATITSKP-----GETSPDSSKIE- 617
Qy      845 EEKEKVKKKKKOEKEETMQATWVKTTPPKHQVQKQGEAYRVYGQWMSIKTHYV 904
Db      618 EKPDKY--SEVSDDMTEHITADKGTOTFLNINHEHDM-----YGG-----YLF 663
Qy      905 RLVYKLPGLNTVNYRKSLGKTNNMDENDESDRKCSRPKIKIEP---SEKDEVK 960
Db      664 R-----PGDTGIS--RPKVQAPCTNRLSNMVCPEAVLKGGKKLVLPDPYEISDETGT 717
Qy      961 GSDAAGADQNEMDISKITEKKDDVYKELLDSDSKPCKEERPEVDDMKTESH----- 1014
Db      718 LSDSEETSPSAEMOSETSEAGPTTIKTSGETERTQSSSPSEPTSDRKMKKKLDT 777
Qy      1015 -----VNCQESSQVDVNVNSEGFLRT-----SYKKTKSKLDGLLERIKO 1057
Db      778 RRRKLADBDLSDPWFYSIDGNELVATGKPIIKHKFYTSANRMPKSLDPTGRNMAE 837
Qy      1058 FYLEEKQRL-----EKILEGKIGKIGTSTNSSKNLSSESPVITAKEGCOSDSMRQ 1110
Db      838 IPMEISRLAEQVAHEEYKIRRRQVLEAVEAASKILNVYDTJEEBE-----IEBE 891
Qy      1111 QSPNANNQDPEDLIQGCSSDSSVLRMSDPSTHT---NKLTPKRVNLDVSTIRSETKC 1166
Db      892 ETPE-----EEVVK--VASPTAPATENPTTSTAFEBEGVAMKETPIEDIFF--DPDEPC 942
Qy      1167 PRONSIENDI-----EKEYSDLASRGOEPTKSTKGNDFIDQSKLASADDITGLIC 1218
Db      943 SSGAQAQRELIIETRGKKEQOIIEDSLQNNKPS--SKT-----VKSSEREA----- 987
Qy      1219 KKKPLIQESDPTVSSSKSALHSVPKSTNDRDAPLSLRANDFECKLACDSESNSTLEN 1278
Db      988 ---QEPRIEKDEMESEQKQKADANPVEYDKESASSESDDKSDFE-----DETDA 1036
Qy      1279 SSGTVSIOSSSEDMIVQVNSNESISQFRT---EDDVLEFLKCELYSGESTGNC 1332
Db      1037 QSKTYAKISLKHRTY--SDEET--EDFTKFEQEVATADAMKIKRTIGEYVSTEL--- 1088
Qy      1333 EDRLPVKGTGANGKPSQOKLEERP-----VNKCSDDIKLKNTJDKKNNERESEK-- 1384
Db      1089 -----KLVAQQPAVYDDEVLAGFCVARNNDQFSITIKETGKRTNNKPNPDESAY 1135
Qy      1385 KQORTSTQIUNGK-----DNKPKIYLGECLEKISERSVVSGVNEPVANNINKIIPEN 1437
Db      1136 KNFRRES-FAAKHRPVPRKLPNTIMERKYI--ERAHNVKYKHV--DWEPHMKILIALAQRO 1190
Qy      1438 DIKSILTVKS---AIRPFLNGVIMEDFERNSSSEKSKILLSSSDAEGVYRQSLTLPST 1494
Db      1191 QIAATCANLSQPVYVTPREHABOVOLLNLPNSILRPL-----NNPQALTLHLKAO 1243

```



```

QY 1495 KESDSTQTTTPASCPESNSVNOVEDMEIETSEKKVKTSSPI-----TSEESNLNDF 1548
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1244 QQAIOQORAAQKQOTOKELARQAEQARVELAKRRIAQDAEKALRQKEQMSNVS--- 1300
QY 1549 IDENGELPINKENNVNGESKRKYITE--VTMTSTVAATESKVTI-KVEGDKQ-----T 1599
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1301 ----GIVSSDQNAQSSNAQOTGLIENQTTTJNSDSLIRPNTLADNSHLGESSQIPVIES 1356
QY 1600 VVSTENCASST-----VTTTTTIVTKLSTPSG----- 1628
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1357 IQSSTSEALKESSENYKDMPLITPASTVSSKSSAPATRRPRPCSSYDRPSSPSVIRERL 1416
QY 1629 GS-----VDIISVKE--QSKTVVTTVTDSLTT---G 1656
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1417 GSDGALINRPNCNDKSRSPISRAPEYETVRINDHGONETILAGNITHYETILL 1476
QY 1657 GTLV---TSMTVSKEYSTRDKVKLMKFSRPKTRSGTALPSYRKFTVKSTKSTFVLND 1713
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1477 GTSIGQDSTIRYDQEGSTQYID--KTIDDNKNGKGTNVEDQSNVLK-----LREN 1526
QY 1714 DLKLLAKGSGIREVPTNYNAKRALDIWYPSRPRTGITWKRRLQTVKSLAGVSL---- 1769
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1527 DLNR-----EMLRAN-----RHPSTMLAMGNLSINERH 1556
QY 1770 -MLRLMASLRWMDMAKVPBGSGSTRTETSEITTTETILIKRRDVGPGYIRFEYCIRKI 1828
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1557 NKVOQVLAISELODLIARHISGA-----VSQTVQVEY-----YQEG----- 1592
QY 1829 IC-----PIG-----VPETPKETP-----TPQRKGLRS----- 1851
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1593 VCAGTSDAIGETDEDDVEEPEFTVDQLAKKILKQROGLSEDSDEDDVDVYDND 1652
QY 1852 -SALR--PRRPETPKQGT-----PVLIETWAEELELMELRAAEVVEKEKAQAVE 1900
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1653 GSIVIRRAPHKKRETRKKNIIVPNIPKIRKIV-----DKKIEEMERA--- 1695
QY 1901 QQAKKRLQEQKPTVIATSTSTSTSTSTSPAKVM-----VAPISGVSYTTGKMWLT 1955
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1696 -KRRARKSKQKMASIRIAV--PQKPTQGFATPQGVNPGPKHSAASAAARATPRPKKAKMS 1752
QY 1956 KVQSPATVTPQON--KNFHQTFATWAKQSGNSGVVOVOQKVLGITPSSF--GTSQOFT 2011
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1753 NVSPPTKITATPVNVAAGFHQ-----NQOOLYSDMAQ-----PQSPRTITTPQGT 1798
QY 2012 SPQPRATVTRIPNTSGSGTSSGOVITGROIAPGKTVIRTPLOQSTLKKIIRPVNV 2071
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1799 GSAFOAOTPOSHLAQLQGFVNGAMQ--QAPQ-QQGM--YTAALQLOAMQAAVAQAQAA 1852
QY 2072 Q-----FGAPQVWTOILIRGQPVSTAVSAPNTVSTPQOKSLTSATSTSIQSSASQPPR 2126
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1853 QAAVAEAAVAQVAAQVAAQARAARPOQLVQROVYGH--PGQVNVIMPRAMMLN-QGNPQAVN 1910
QY 2127 PQGQVAKLMAQLTOLTOGHGNOGLTVIQQGQGTGGLQILPQGVTVLIRGREQOLMDA 2186
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1911 PAQAQMDDERKKMEV-----NAVYHLMQSGQ-----PPPTNOELFQV 1949
QY 2187 AMNGTVQRLPLPLATTTATTTTSTTTTSTAAGTGEOROSKLSRPMQVNOHOKTILPPAQ 2246
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1950 QF--AAQADLRSAADAADAQAQAQOMTMRQAQAEVAQAQAMKQEOAR-----AQ 2000
QY 2247 SSSVGAPOAPOTAPSPARPOQTOPSPAPQEPVOTOPREVOTOTVSSHPSSEAPQTHAQ 2306
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2001 AAKKEARALKAEFEAKAKAKQAQAEARAKAEQEMRVR-----QAQAQOTQ 2045
QY 2307 SSKPQVAAQSPQSPNOVQSPVAVQSPQSTRIRPSTPSQLSPQOQSOVOTTTQOPIDIP 2366
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2046 AAQAQASQSAHAQNOAQOTQATLVEIORMIQ-----SCQPLSMQ 2083
QY 2367 HTSLQIPSQQPSQPSQVQVSTQTLSSGTLNVYS-----VSSPSRPOQIQIQPQ--- 2417
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2084 MOQLQMSQVQOMQHAQVQVQOMQOMQOLOMQFAARMQOGTEPRPAVSQAQVQOGMPAGI 2143

```

```

QY 2418 QVIAVPOLOQOVVLSQIQSQVVAQI-QAQQSGVPPQIKLQPLTQIQSSAVQTHQIONV 2476
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2144 QGMPTGNGMQDGLGILPGNQMLPQQAQSQQTQQAQQQQLFMQLQIQQQQLMQ--HQLOQQ 2202
QY 2477 VTVQAASVQEQDQRLQDQDQKQKQOQIEIVNWTNPSTKLLIVEII-QQVYMKHNAV 2535
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2203 IQMQQHQQHQOQOQOQOQIQMQ-QOQOQLOQGLVYVNSS-AWLOQQAOLAAQOQOQVVOQNL 2259
QY 2536 EHLKQKKSMP 2546
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2260 --MVPVSAQTP 2268

RESULT 5
T15348
hypoethetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15348
R:Gatung, S.
submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: EMBL:050071; NID:g1208871; PID:g1208877; PIDN:AA93447.1; CESP:B0
A:Gene: CESP: B0350.1
A:Introns: 48/1; 5039/3; 5116/3

Query Match 3.3%; Score 490; DB 2; Length 5170;
Best Local Similarity 17.3%; Pred. No. 4,7e-10;
Matches 640; Conservative 540; Mismatches 125; Indels 1274; Gaps 162;

QY 1 MVSSEEEEDGDAEETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFRSHST 60
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 566 LVSEHEHHPACDSERL--SEPAQSPPEVETHASQFSNLVETTTTGA--VTRFYDEDE 621
QY 61 YSSTPGKRRKRVIRPRSPILFEKDIPPLEPKS-----SEDLMPNHEIMNVIAIYELVN 116
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 622 EQASSRATKERIE--QSPVASERSIVSTEHRSQSPQSEOSLPTSE-----KN 668
QY 117 FGVTLRSLPRPFDFCAATVSOECITLMAEMHVLLKAVIREEDTSNTTF---GPADLK 172
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 669 VHTVETTTVTRCFEPIASF-----LEHASREFEQGSNDPRRSVSPVDQA 715
QY 173 DSVN-----STLYFDGMTVPEVLRVYCESDKYHNHVLPRYQEAEDY 213
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 716 EQVDVHPSPAPSSHAESQGVPPSPHHIVETTTTIVTREF--QDEEYPR--DESPAETP 771
QY 214 PYGPVENKIVLQFLVDQFLITINARELMSQVIOYDDHCRCRHLGLDLCCEQSAVY 273
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 772 PIPSSQOSEPHIVKETTITTTVTRF-----LYDE-----PRKGVNTFSPADSS-- 816
QY 274 HLEGVKPRLEVEPEDEMOCEVCVAHKVPVTDCAVETIQKKPYIRREPLIYDSSRRKMYF 333
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 817 HAESR---QVP-----SPVYSIQEYH----- 837
QY 334 LNRLLIEDTENENKRIWYSTKVQALIELDCLDKDYAEALCKILEEMREELIR-HM 392
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 838 ----VETTTTNVTSNIDDEDNVSSB--DPATQHFQOSE-----TSVHRSHP 881
QY 393 DITEDLTNKAQS-----NKSFLA---ANEELIESITAKG-----DIDNVMS 433
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 882 DSVSESDGEGISKVLGFAKAKAGVAGVAAVLAALAAVGAKAAYDAFEKDEDDDETSHS 941
QY 434 PEE--TEKQKNETENDSKD-----EKNKEEFEDQ----- 461
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 942 PESPVPPEYQEQYQDSSAQSSHTDFEHNMPESTIEHKEETEEFDHSHPSVLSKEKEH 1001
QY 462 -----SLEKSDSDKTPDDDPQGGK-----SE 482

```


Db 1002 QVTSFTTTTTVYREXNDPEDEQEHQPHSPASSHTAEHPHIVTTTTVTREFOE 1061
QY 483 EPEYV-----GDKGNSVSNLGDN-----TTNATSEETSPSE--GRSPVGLSETPPOS 528
Db 1062 EPEBELEYKQEDNSKRSPSSHQENLVTETTTTTVTSEYDEPHHEEOETGKSSAPSS 1121
QY 529 SNAEKRVASSELPODVEEPNKTCESSTNTSAVTTTSIQPN-LENSNS-----SELSNS 580
Db 1122 HSESEROY---LESPPASDPRIHVMETTTTTTTTVTRQFHODESESRSDPNRDVYEAQIHSS 1178
QY 581 QSESAAADDPENGERSHTPPVSIQEEIV--GDFTSEKSTGELSESPGAKGASGSTRII 638
Db 1179 NIESHQFSEKDDDDSGRSPSPKSEDEPVKHESYKFEISTTE-----TRER 1224
QY 639 TRLRNDPSKLISQOQVAAAHANKLFKEGKEVLVYNSQGEISRLSTKKEVIM-KGNI 697
Db 1225 DR-----PELDSERLSEPA-----OSPMEVSPITTEENIVRSPSSV 1260
QY 698 NMYFKLGQEGKY-RVYNHOYSTNSFALNKHNRHEDHDKRHLAHKCLTPAGEFKWNGSV 756
Db 1261 KSEY--GSEGHVPSVLETGTTTTVTREFYDDDDQDEHENOQOSELSAIPTEEEHEHGS- 1317
QY 757 HSGKVLITLRLTLTQLENNIPSSF-----FHPNMAHSHRANWTKA-----VQMG 802
Db 1318 HLEKETTTTT---TVTRREFYDEPENVELODDQFSPAPSPSHVESGIIHASESPVAQOQOEIP 1374
QY 803 KPREF-----ALALAILBCAVK 819
Db 1375 QTRREFHEDSPAQVHHEEYESHVLTQAPRLTEQOHQSPSGDESGEGISGVILGFPARK 1434
QY 820 P-----VYMLPIWREFLG---HTRLAHMTSIE-REEKEVYKKEKKOEETEMOQATW 868
Db 1435 AGWVAGVAAAPVALAAGAKAAYDALKKDDDEQEBERSILRQRSIDSPPHASQS-- 1492
QY 869 VKTTPVKAQVAKQK---GEETRVYTGKMSVTKTHVRFVY---KLPG--NTVWNTR 919
Db 1493 ---QIEBEHEREEESPVPBEKHNVHTTTTTVTTRFDEHPELVSOEIEGEGSPADSE 1549
QY 920 KSL-----EGTKNMDENDSDKRCSPKIKIEPSEKDEVGSADAKAD---Q 970
Db 1550 KSLPHVETTTTTVTREDKNDSE--SPVPSKEQEPITVSREV--YETAEBDEPHNY 1605
QY 971 NEMDISKITEKKDQVKELLDSDPKCKEEMEV-----DDDKMTE-- 1012
Db 1606 TETTTTTVYKEVIDDQEM--GDDEPKQESPQVVETTTTTMSRVEDDDDBTRSEAGD 1663
QY 1013 SHV-----NCOESSQVDVY-----NVSEGFHLRTSKKKTKSS 1045
Db 1664 SHITETTKTTTVVREPHGQPEJETEDEVELLPKIEEDNVSEYSESTSVSREVRPD 1723
QY 1046 KIDGLLEERRI-----KOFILKEQRLKIK 1070
Db 1724 E-PHIETTTTTVTREYINPEPEYDDOKDAAPISFQEHODDQASDHDRSPVE 1782
QY 1071 LREGIGICKTST-----NSKNLSESPVTITKAKE--GCQSDSMROEOPNA- 1115
Db 1783 SEKSVNHTTETTTTTVTROLYDDDEASEIRGESPVATEEHEHNSIKSDDESQHPSVY 1842
QY 1116 -----NNDQPEDLIQCCSOSDSSVLRMSDP-----SH-----TTN 1145
Db 1843 ETTTTTTVTRREFYDDQDE--LQREDHTQSEERRSSIPTEETEHEHSHLIKETTTTTVTTR 1900
QY 1146 KLYPKORVLDDVSIRSPERKCKPRKONSIENDIEKVDLSRQGEPRKSKTKGNDPFILOS 1205
Db 1901 EYDEPEPNE--KLDSQPSLSPSSHVESEIYVESPVAKQOEIPQTR-----FHEOS 1952
QY 1206 KLAS--ADDITGLCKNKKPLIOEESDPTVSSSKALSHSVKSTNDRDAPL-GRAND 1261
Db 1953 PAAQVHDEYEHQVPTBAPRLITEQO-----HQ--PESEBESDGEFGSKVLG 1999
QY 1262 FEGKLGCC-----DSNSNTL--ENNSDPTVSID 1287

Db 2000 FAKKAGVAGVAAAPVALAAGAKAAYDALKKEDDEDEQEBERSILIREBSPDSPHASE 2059
QY 1288 SSEEDMIVONSNESISEOFRTREODVEVLEPLKCELVSGS-'GNCE----- 1333
Db 2060 QSOIEKEHKRFESPPVPSEKHNDHQSALLQOESVQIENKESRFFNDESERGVSKDHYTED 2119
QY 1334 DRLPVKTEANGKPPSQ--OKLEER---PVNKCSQOI-KLAKTT--DKKNENRESEK 1384
Db 2120 DQESLKPKEGCAFSQFTSEKQEDRSDSPIHQKEDISQFONESPPEDVKSQOPHDEK 2179
QY 1385 KG-QRTSTQIINKDKKPKIYLGEBCLKEISERVVSGVNEPVRVNNIKTIPNDIKSLT 1443
Db 2180 PDLERGSYSSGSPKSP-----GGSITGLDEKALSGVQEPDR-----PENFAESHB 2228
QY 1444 VKESAIRPFGINDVIMEDFERNRSSETKSHL-----LSSDAEENYNDSETLTPSTRES-- 1497
Db 2229 KTEAT-----SDENLFESDKYAPARSPVPSDESSNRVLETITTTTTVTRHRE 2274
QY 1498 -----DSIQTPPSASCPESSVNOVEDMEIETSEVKVYTSPTSEEF--SNLSN 1546
Db 2275 PEDDHSYVVESEQEYSSSGSPVPSEKSVDRV---LETTTTVTTRHFEADEIPIVIES 2330
QY 1547 DFLDENGRLPKNNENNGESKRTVITEVTMTSYATE-----SKTVIKVEKDK 1597
Db 2331 SHDDQASAPVSEEDVHEQ-----IQTTTTVTYKHEFVPDDEIDSEHNMESOKVAS 2383
QY 1598 QIVVSTENCARKSTVTTTTVTYTK-----LSPTSGSVDI--SVKEOSK 1641
Db 2384 GSPVPSSEBDSKRYETTTTTVTTRHFEPEBDHSPPVQOYEYASASSPVPSEKSERVIE 2443
QY 1642 TVVTTTTV-----DSLTTTGGTLVTSMTVSK-----YSTRDVKLMEF 1680
Db 2444 TTTTTVTTRHFEDEDHILGQGESDSQIPSEISITSENMORETSSPVSQNRDEEVLPA 2503
QY 1681 SRPK-----TSGALPSTYRKFTVTKSTKKSIFVLPNDDLKILARKGGRREPVYRN 1731
Db 2504 IAPYKQPTXGRVSDHAPSPA-----PSAESPQIQAQKQESQEAHSLQEFQOQSSVS 2557
QY 1732 YNAKPALDIW-----PYSPRPFTGJTWRYRLQTVKSLGAVSLMLRL----- 1774
Db 2558 HEDSPAAQVFNHESDHSDSVPSDRAPRLTEQOHQESGESGEBFGSVLGFARKAGM 2617
QY 1775 -----WASLRWD-----DMAKVPBGSGSTR-----TETSETE 1802
Db 2618 VAGGVAAAPVALAAGAKAAYDALKKDDDEDENPDQEKLLPSPROVLYNVPESDSE 2677
QY 1803 ITTTEILIKRDVQPYGIRFECYIKRILICPIGVETKEP---TPQKGLRSSALPRKR 1859
Db 2678 ISEIE-LEVTSPSPTEKSESQCYETVTRTTTVREYLDDQSVTRSRSPSEHDISQYAP 2736
QY 1860 ETPKOTGPVIEETVVA-----EELEWEIRAFARVREKKAQAAVEQOAKR 1906
Db 2737 ESPVEDDPVVEKTTTVYIROYNDEPPOLEIEOIIPREYVYLKREYVSPBEDDEQH---Y 2793
QY 1907 LEOQKPTVIATST-----TSPTSSTTSTIS--PAQKVAVAPISGVTGT----- 1949
Db 2794 IETKTTTTITKEVHNVAVEEDVQIISPVSHSETSVEKQLPADQOLDEVTSTTATVYTRER 2853
QY 1950 -----KMLLTQVGSPIATY---FOQKNFHO----- 1973
Db 2854 YEEPEVRPPSGSEADESHAPKYMETT---TTTVYTREREVESEDEHDOQSOVORDESP 2910
QY 1974 -----TFATWYK-----QOQNSGVVQVQOKYLGIIIPSSGTSGSQOFTSFOPR 2916
Db 2911 APSEDSVKHVIETTTTTVTEERYEPEDSHSPSPSDDVHGKVTTTTTVTYVHEHFEPE 2970
QY 2017 TATVTRPNTSGSGGTTSNSQVITGPQIRPGMTVITPRLQ-----OSTLGAAILRFPV 2069
Db 2971 -----DRPDSBHVESERIVASGSPV---PESEBDSRLETTTTTTTTVTRHREH 3013
QY 2070 MQQPGAPQOVMTO--IIRQOPSTAVSAPNVVST-----PGQSLTSATSTSN 2117
Db 3014 ELEDDEHVESQEYASGSVPVPSEKSVERVLETNTNTVTRHFEHEDDIPIVETSHD 3073

QY	2118	QSSASQRRPQ--OGOVKLTMAQLOLTLOSHGNQGLJTVNIOQOQGTQOQLOLIPQVTV	2175
Db	3074	DPAAVSVEEDVHQIO-TTTTTIVTVTRH-----VWDEETDSGMDELKYSSE	3124
QY	2176	IPGPOQLOMAQMPNGVORFLEPLATVTASTTTTTVSTTAAGTGEOBOSKLSPOMQ	2235
Db	3125	SPVPSEE-----DSKVIETTTTTVS-----	3146
QY	2236	VHODKLTLPAAOSSVGPAAKAPQTAQPSANRPOQTOPSPAPQEVOTQPEVOTQTVS--	2293
Db	3147	--REHFEEDEDSHV---VESOEYSASGSP-----VPSEKSEVERVETTTTTTVTRE	3193
QY	2294	-----SHVSEADQPTHAQSSKQVYAAQSPQSNQOGSPVNVQSFQSTRIRPISQOLS	2347
Db	3194	HFTDEEDTIPESKTSHTDDGTQD-----HVPQSPV-----PSE-3229	
QY	2348	PGQOSQOVTTTTSQPRTPIQPTSLQIPSOQPOPOQOVQOSTQTLSSGOTLQNVSSPSR	2407
Db	3230	DDHVTHEQIIMDDPIIDQOIVESHKYSK---SSIPSEDSHTVIT--ETWSTPITS---	3260
QY	2408	POLQIQOPQOVIYAVPOLQOQOVOLQIOSQVVAQIQAOQSGVPQOIKLOLPTQIOQSSA	2467
Db	3281	-----ERYPDEV-----EKDVESADEI-----DSSAQYKXSESPVQJESKSL	3320
QY	2468	VQTHQIQM-----	
Db	3321	LAKQOQESDESDGSGSVLGPRAKKAGNVAGCVAVAPVALAAGAAAYADALKKDDDE	3380
QY	2495	RDOQOK-----KQOQIEINVTPSKLTIKV-----ITOK	2525
Db	3381	EDQEDKEPLIGFHKDDPIISQDEPSAVDSQEPATYPEREEDKVTITDSABSVQD	3440
QY	2526	Q--VVMKHNAYIELKOKKSKTPAERENORMIVCNOVKYITLKDIXKEQOAKKR--	2560
Db	3441	EPKIVFPVDSIPRHSN-----DREEFESIV--KSESPYIVESTDYAQTSAPRIS	3491
QY	2581	-----KREESV-----	
Db	3492	PVHSDAGDASSFKPESVTEGDEKMAIPETSETDAPYIDSEYEFNNRMDQRISSPAHS	3551
QY	2592	-----KONATKLALFHKQELRAETIKKRALDKLOI-----	2626
Db	3552	DEDEBNDAVEIDSFYHSDQONNE--EDPSIVSESEYISSGHGSPREFEDSTTTVYINV	3609
QY	2627	-----EVOBELKRDIKKEKQDLQLOAQATAVAACP--VTBVLAPRAPPESP	2674
Db	3610	HHEPAIIPREPVDE-----ELQOERSITE--SEKYTSPLPPTSVTTVEHVERAE--	3659
QY	2675	PPPGVQHTGLSLPRL--PYASOKKRRBEK--DSSSKSKKKMIISTTSFKTKDKIKLY	2730
Db	3660	-----IHKRTTSPVITVTVSSSEHLDRQOEPYVESESEYTNASPLGEBRESGSP-	3712
QY	2731	CICKTPYDESKFYIGCDRCQNMVHGRQVQILOSEALIDEYVCPOQOSTEPAM-----T	2784
Db	3713	-----PREEDSH-----VIESH-----EYISSVPYS--EDSVKHVIEKT	3745
QY	2785	VLTPLTEKDEGLKRVLRSLQAHKMAPPLEYVDPNDADYGVIV-----E	2831
Db	3746	TTTTTVTEERYE-----PEDSH-----SPVPSD--DVHGFKVKTTTTTTVTTHEHE	3789
QY	2832	PMDLATMEENQRRKY-----EKLTGFVADMKIPIDNCR--YINPSD	2871
Db	3790	PEDTIOSEHVHVESERYASGVPVSESNRSNRVETTTTTTVTVREHEPBD	3838

```

RESULT 6
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favello, A.; Vaudin, M.

```

submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-References: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GND0021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1, 48/2, 84/1, 196/3, 303/1, 381/1, 586/1, 605/1, 1175/3, 1207/1, 1409/
3504/1

Query Match	3.1%	Score 463.5	DB 2	Length 3507
Best Local Similarity	19.0%	Pred. No. 2.8e+09		
Matches 444	Conservative 28%	Mismatches 857	Indels 755	Gaps 92

OY	419	ESIRAKKDDIDNNVSPETLEOKNHEITENDSKDAEKNE---	EFEODSLKDSDDKPTDD	475
Dd	997	ESSSTSTDCKSGGK----	PADKSDNNNVKPDGNLDTTDEEDSETSEGGQNP--	1046
OY	476	PEOGKSE-EPTVEVDKGNSVANLGDNTTAATSEETSPEBGRPVCLSETPDSMAEK	534	
Dd	1049	--KGKSNQPP--GDKOSEVK-----	KPISEVDP-----	CNLGTGKGSNWPLK 1086
OY	535	KVASLPDVBEPBNKTCESSNTSATTT-SIOPLENSNSSELNCSQE----	SAKA 587	
Dd	1090	-----PDLDEEGSGLITTSGGKNSTFPHGTKEPLPKTPEDKSSFPOLGLEISAGK	1144	
OY	588	ADDPENGREHRETVYSIQEIVGFJTFSEKGTGELSEBPAGKASASTIITRLRPDSC	647	
Dd	1144	KPEPDG----TSKEGTELIMESTPGSTTSDSVGLEISDULT-ARKKHVEIE	1197	
OY	648	LSQLSOOVAALAAHEANKLF-----	BKEVLUV-----	NSOGEISRL 685
Dd	1198	GSGTGDEHTATTTHDVSKSTIKPRVEVDGDNGETSVDCKPTTPAPPTSSSAESTSRH	1257	
OY	686	STKREVIMKGINNYFKLOGEGKYRVYHNOYSTNSEPALNKHOHREDHKRHLAHKFILT	745	
Dd	1258	PTTSEASPEG-----	SGEAGVESPDSGESTSA-----	1288
OY	746	PAGEFKMNGSVHSKVLTISTLRITTOLENNIPSSFPHBNASHANIKAVOMCSPR	805	
Dd	1289	PDG-VSPPTSATAEVEVPTTSASSTPDAVESSGJPST-----	SKPT 1327	
OY	806	EFALATALBECAVVRVMPLRMREFLGHTLYHMSTIEREEKKKYKKKKOEBEETMQ	865	
Dd	1328	AEPLETTAPSTEY-----	TSPESGTEESTLLPTBEGSGSTSS 1366	
OY	866	ATWKTFFPYVHQWMKOGBEEYRVLTGYGSMWISKTHYRVFLPGNTNVYRKSLGT	925	
Dd	1367	APUYERATVLRPOQRNRKREPTKDT-----	FALPTTTGAQOAN-DSEVENT 1412	
OY	926	K-KNMDE-NMDESDRKCSBPKKIKLEPDSEKDEYKS----	DAKAGDON-----	971
Dd	1413	KCTSSDECDL---ALCERRTGVCOREPEFEBAGPKKSCVDDECATGDHNCHESACQ	1468	
OY	972	-----EMDISKTTEKKDDOVKELDSDS-----	994	
Dd	1469	NVYGVAFCPTGRFRKADDSOCODIECTEHNSTCCGANAKCVNKPITYSCENGFLGD	1528	
OY	995	-----DKPCRKEPREVEDDMKKTESHVNCQSS--	QVDVVANS-BEFLHRTJKYKKT 1042	
Dd	1529	GYGCVPVPTTKPC-----	DSTOSSKSH-CSESNNSCSEVDTVDGVSVECKECMGYKSG 1579	
OY	1043	KSKR-----	LGLLERRIKO-----	FTLEKO-----RL 1066
Dd	1580	KVCEDIINECYAEKAPCSLIANCYNMNNGTBSJCSCOGYRGDCFMCTDINECDERHCPHPA	1639	
OY	1067	EKILLEG-----GIKGIG-KTSTNSSKNISE-----	SPITTAKEG-----	1101


```

Db 1640 ECTNLEGSFKCECHSGPEGDGIKCTNPLERSCDEYKPCGRVDHVSCLSVRIYNGSLSS 1699
QY 1102 -CQSD-SMROGSPNA-----NDQPDLL----- 1124
Db 1700 VCECEGPFREKSNCSVDIDECESRNNDPASAVCVNTGSGYRCGAEYEGGVCT 1759
QY 1125 -----QGQSDSSVLRM-----SDPSHTTN----- 1145
Db 1760 DIDEORGNAGCDMSAMCINRMGSCGCKMAGYTGDAFCIKIEEPKSDKTADEWSR 1819
QY 1146 --KLYPRDRLVDVSIKRPETKCPKONSIENDIEE-KVSDLASRGOF-----PT 1191
Db 1820 LCELEKKQCVDEEYVPGGACLPGHHPINGTQSLQISLCLQKXNDCKNHACIDIHDP 1879
QY 1192 KSKTKNDPFI-----DQSKLASADIGITLIC-----KMKPLI 1225
Db 1880 SHRCSPDGFIDGMICDDVDDECNNAGMCDDENTKCENTIGSFNCVCLGEGFKKVEDECVY 1939
QY 1226 QESDITVSSSKSALHSVPKSTNDRDATPLSRAMPEGLKGDSESNSTLENSDITVST 1285
Db 1940 DEKKOP--NKEKIEIDENSSSSGQEKPTTKGI--VSSATSSSESTTAEPHVTTSSIS 1995
QY 1286 ODSSEDMIVONSNEISIEQFRTREODVEYLELKCGLVSGESTGNCEDRLPVKGEANG 1345
Db 1996 STSTDMTSSKSPENVMTMSSEPEVST-----SKSTTASETTIVSSTPSESS 2044
QY 1346 KPSQOKKLEERP--VNKSDQIKLNTDKKNENRESEKQORTSTQINGKDNPKI 1403
Db 2045 --SSAPLTSSPATTEVTTSESSVKSTPKESSESEIVKLSSK----- 2087
QY 1404 YLKGELKEISESRVYSGNVEPV--NNINKITPENDIKSLYKESAIRPFGINDVIMED 1461
Db 2088 -----PEVTESSVSSSPSTPTTQSQSVTVPETSKSVLSSSEAPVSTPTEV----- 2136
QY 1462 FNERNSSETKSHLLSSDAGNRRDLETLPTKESDSTQTT----- 1503
Db 2137 --HTSEETKPLSLASSTT---GDTNSTTPRSSLASVAKSTAPGCTASAVPVLSL 2189
QY 1504 TPASCPESNSVNOVDEMEIETSEVKKVTSPTTSEESNLNDFTDENGLPINKNNVNA 1563
Db 2190 SPVSGQSPKTFPATSESTVQASSETSSGTSVKSTSEPSHVTKLSTSS---NPSSVAV 2245
QY 1564 GEEKRTVITETVMTSTYATEKTIYKVEKGDKQTVVSTENCAKST---VTTTTTV 1619
Db 2246 VTSKSTPTVPBSTEQPTSTPBGOSLTPMNSNEVLTTSEPVVLSLSPDVOSQSTTP 2305
QY 1620 TKLS-----TSTGGSVDIISVKEQSKTVVTTVDSLTGCTGLVTGWSKEKSTBD 1673
Db 2306 NNLSSESTVETPKTSEVNLNSEEPTTEAPTLSPDILSTTNNLSQSTV---STED 2361
QY 1674 KYKLMKFSRPKTRSGTALPSYRKFYTKSTKSIIFVLPNDLKLARKGIRE-VPEFNY 1732
Db 2362 RSEISENSEKPTSAPELVTSVTHVASSPD---VPRESSPDDLUGSTSTNIP--EA 2415
QY 1733 NAKPALDIPYP-----SPRPFGITWRRLQTVSLAGVSLMLRLMLASLRWDMANAV 1787
Db 2416 SSKQRTSSPTPTDPTTASSEPKTSTMSPDLSSTSNVLSSES-----ST 2458
QY 1788 PRGGGSTRTESETE---ITTEILIKRDVGYGIRFEYCIRKIICPTGVPETPKETPT 1843
Db 2459 TEBSSEKSVSSSTBGISVVTSTERSK-----VPS----- 2489
QY 1844 PORKGRSALRPKRPETPKQGTGVLIETVAEELE-----LMEIRAFAEVY 1891
Db 2490 -----TISVLEEDLTKT--TSPFILEETTTASETSEPLEDLSLTVSVRIHELTTSENV 2542
QY 1892 EKEKAQVQOAKKRLQOKPVIATSTTSPTSSTTSISPAOKVAVADISGSVTTGTVM 1951
Db 2543 PKE-SESTTSSSESKPQSEPAGITLSTIVVPTSSVSLITASE-----IEATISNTPF 2594
QY 1952 VLTTKVSGPATVTFQONKHFQTFATWVKOGOSNGVVOQKVLGIIIPSSGTSCQCTT 2011
Db 2595 -----KQGRK-----PITTSPLSKLVKST--TSPSTIVT 2619

```

```

QY 2012 SFQPRATVTVIRPNTSGSGCTTNSOYITGPQIRPGMTVIRPLQOSTLKALIRPVWY 2071
Db 2620 SSEPSESTKR-----TIVSTTVST-----TPTTEETTSSESLILT---A 2655
QY 2072 QPAPQOVMQOILRG--QPVSTAVSAPNTYSSRPGQKSLTSATSTNIOSSASOPRPOQ 2129
Db 2656 APSKPTSESTSESEAPPTPKTSETKPSNVSS-----TSRSTEVETSTSO----- 2702
QY 2130 GQVKLMAQLTQLTQGHGNGGLTVVIOGCGQTTGOLQILPGCVTVLPCGQQLMOAAMP 2189
Db 2703 -----SGSLF-----SSTMS 2712
QY 2190 NGTVQRFLEPLTATATATSTTTTV-STTAAGTGGOROSKLS---PQOVHODKTLPPA 2245
Db 2713 STSEPEETNAPAVTVSSEASSSTLEENSSSTSPSSSEASVLSLSPESITSBVTV--- 2769
QY 2246 QSSSVGPAKAPQPTAPQAPARPOQOPQOSTAOQEVOTQPEVOTIVSSVPSSEAPTHA 2305
Db 2770 --SSRAPAEI--TMSSESHREISTVSSPESEPEILPLSTVSPNVYASSIPSE----- 2818
QY 2306 QSSKPVAAQSOCP--OSNVQOGSEPVROQSPQRTIRPSTPSOL-----SPQOGSOVOTTS 2359
Db 2819 -----EPILSSVYSSSTPRV-----RLITGTPDDLIVSVTVPSHGKRRQNTA 2861
QY 2360 QPIPIQPHTSLOIPSOQOPQOSPOVOSSTQTLSSGOTLMOVSVSPSRPQLOIQOPQ 2417
Db 2862 SSVPSNSTSTILPSSLSLTPQPP-PTTTTAKPATTSGKRGPPSIQPAEMFTTPAP 2918

RESULT 7
A35085
trithorax protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999
C:Accession: A35085; A38240
R:Maizo, A.M.; Huang, D.H.; Wozer, B.A.; David, I.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 2112-2116, 1990
A:Title: The trithorax gene, a trans-acting regulator of the bithorax complex in Dros
A:Reference number: A35085; MUID:90192757; PMID:2107543
A:Accession: A35085
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3759 <MA2>
R:Maizo, A.M.
submitted to GenBank, January 1990
A:Reference number: A38240
A:Accession: A38240
A:Molecule type: mRNA
A:Residues: 1-2361, 'Y', 2363-2397, 'N', 2399-2405, 'N', 2407-2411, 'N', 2413-3759 <MA2>
A:Cross-References: GB:M31617; NID:9158817; PID:9158818
C:Genetics:
A:Gene: FlyBase:trx
A:Cross-References: FlyBase:Fbgn0003862
C:Keywords: Drosophila trithorax protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match
Best Local Similarity 3.1%; Score 462.5; DB 2; Length 3759;
Matches 627; Conservative 422; Mismatches 1151; Indels 1157; Gaps 156;

QY 5 EEEEDGDAAEETQDSEDD---EEDEMEDDDSDVPEEMEDDDDDASCTSSFSRSHST 60
Db 538 DEDEDEGVTFRRRDSPEDDQNNMEDDDDDDEFAEEDDQNDDEADDEASAKSATERS 597
QY 61 YSS-----TGR--RKPRVHRPSRPILEKDIPLPEPKSSEDIAMVN 101
Db 598 AGADERDPDEKOLVMDSHFVLPKRSRSLRIKPNRKLLEBGA1-----STKKPLSLGD 651
QY 102 EHMNVIAI-----YELNRFGT-----VRLSPFRE 129
Db 652 SKGKNVFGTSSSAGSTASTFSASINLKLKGTFFENFGTLKPNSSAAG1FVLRQPLQO 711

```


QY	130	DFCAALYSQECJTLMAEMHVLKAVLAREDDNSNTFFGADLKDSVNSTLYITDQMTPE	189
Db	712	-----ADNOQATFAAARKACPTSPSPAIKPRANSLATSSFGSLASTSSIV-----TPPS	760
QY	190	VLRYVCE---SDKRYHHVLPEYOAEADYPGVPENKI--VLOPLVDQFTTNJAREELMS	244
Db	761	ACSI--CAVVYSSKVV-----TQARKGVVACDVCRFFSKMTRKKSLSANSTRAN--TS	810
QY	245	EGVIOYDHCRCVCHKLQDLIC-----CEYSAVYHLCVYKPLEYEPEDEMQCEVC--	295
Db	811	SGSOOY-----LOCKNEGSPCSIHSAKSQULNFKRFFKYD-----RCTACWL	852
QY	236	-----VAHK-----VPGVY--DCAVAFQXKKPIYRHEPPIGYDPSRKYFYFLNR	337
Db	853	KCMISFQPLPAHRSRSLAILPGMGGEAAAREKSAELL--SPGSLF-----	900
QY	338	LIEEDTENENEREKIWYSTRKYOLAELIDCLDQDYEA--ELCKILIEEMKEEIRHRHMT	395
Db	901	-----TSTAASSPEV--VASISVK-----WKSSGDSLTALITKPRPLANNVT	943
QY	396	EDLTNKAQS---NKSFY-----AAANEELIESIRAK--KGDIDNVKPEETEKD	441
Db	944	FSGPILRLPALLENPLFLKISNADQKLAAEALISPLTKRKSQKEKEVKESEQSEKLL	1003
QY	442	NETENDSKD---AKNNEEEDDLSKDSDDKPRDDPQQKSE-----EPEYGD	489
Db	1004	SPYAGTKRSCGAAYOEYEOYPOKPEAPQTSITTQPSASNGASHGVPOALAGELNMGD	1063
QY	490	KGNVSANLGDNTJNATSEETSPSEGRPGCLSE--TPDSSNMAEKYASLPDY--	544
Db	1064	TLKQRIDLPRPKXKHCGRASATYLOQ--PLATGEOQOEYEDADMQOELTAAYVPSAIMEP	1122
QY	545	-PEEPNKTCESSTNTSATTTISIQPLENSNSSSELSNQSQESAKAADDPENGERESHTEVS	603
Db	1123	SPEKPTHTVTDENDNCASCSTSPGDESKPSKSGSAQAEVKA-----TA	1168
QY	604	IQEELVGDFTSEKSTGELSPPGAGGASSTIITLRLNPNPSKLSQLKSQVAAAHHA	663
Db	1169	LQKE-----TASASAGSSSAKVT--TKN-----AAAS	1194
QY	664	NKLFEKEGVLEVNVSOGEISRLSTKKEVIKGNINNYFKLQEGKRVYHNOYSTNSFAL	723
Db	1195	NLIYAASKR-----QRNDI--ATSSSVYQSSN-----QYQGRKTRKHROQRI--L	1236
QY	724	NKHOHREDHD-----KRRHLLA-----HKFCLT	745
Db	1237	ISIDFMEWYDPAEVCQTFGLIVETVYAQRALCFLGSTGLDPLIFCAACCCEPYHQYCVQ	1296
QY	746	PAGEFKMNGSVHGS---KVLITSLRITITQLENNIJPSSFPHNMASTRANMI-----	795
Db	1297	DEYNLK-----HGSFEDTTLMGSLLETTVN--ASTGPPSSL--NOJTORLNMCPRCIVC	1347
QY	796	-----KAVOMCSK-----PREFALALILECAVAPVYMLPIW	827
Db	1348	YTCMNSSSKKCKQCKQKNHSHTCLGSKRLGLADRPILCVNCLCKSKSCTTKV--	1401
QY	828	REFLG-----HTRL-----	836
Db	1402	SKFQGNLPMCTGCFKLRRKKNFCPICQRYCDNDPDLKMMEGCQDGOWHSCFEGLSDEQ	1461
QY	837	HRMNSIEEKEEKYKKKEKQOEBEETMQCATWYKTFPKYQHWYKQK--GEIYRVYIGCW	895
Db	1462	YNLSLTPLESIEFLCKCARNESS-----KIAEWRQAVAEERKASLYVL	1509
QY	896	SMWSTHYRYRVPPLPONTWVNYRKSLGCKNNMDENMDESDKPKCSRPKRIKEP---	952
Db	1510	KLSKSRQACALLKLSPRKV-----RCTCGASSNCKQLOPKAL	1548
QY	953	--DSEKDEVKSDAKADONEMDISKITEKKQDYKKEILSDSDKPCKEEPEMEVDMMK	1010
Db	1549	QFSSGSDNGJGSD---GESQNSDDYVEFNDQOQO---QQRNNNMNKP--RYKVL-----PCS	1598
QY	1011	TESHVNCQES--SOVDVNVSEGFHLRTSYKKKTKSSKLDGLLEIRIKQFTPEEKORLEKI	1069

Dd	1599	CQOHSHSQSLSLYD-----KOKINGNSVYSLSEENNYMSGVIO--	1639
Oy	1070	KLEGGIKIGKTSTNNSSKNLSESFVITTRAKEGCQSDSMRBOQSPPANNDO-----PED	11222
Dd	1640	-----SNCELDIDIAK-----ELLSEQPFMEQNETKACTDALEED	1674
Oy	1123	LIOGCS-----QSDBSYLMSDPISHTTKLYPKRVRIDDYSINSPEFKCKXOXSIND	1175
Dd	1675	MFEESOGCNVEDIJDAGVASAVYNHEHSTSQAQRSRGVD-----IPL-----	1717
Oy	1176	IEEKSVDLASRGOEPTSKTNGDNFDIFDDSKLASADIGLICLNK-----EPKI	1225
Dd	1718	--EYDDBFGSGCIK-MKLDIRMCLFCCKXSEGSJGBEALILCGHDCWHTTCAMWSAIV	1774
Oy	1226	QEESDTIYSSSKSAHLSSHVSXKS-----TNDRDAT-----PLSRAMD---	1261
Dd	1775	FEEID-----GSLQWHSAAVARGMRIKCTVCNKGATVGCNVBRSCGEHYHPCARSIDCAF	1830
Oy	1262	-PEGKIGCDSMS---NSTLENSDYTSIQDSSEDMITONSNEJSIEQPRFREDOVEYLE	1317
Dd	1831	LTDMSMCOPAHAKGNALKANGSPSYVYESNFEV-----SRPYVELDRKKR---KILE	1881
Oy	1318	PLCKELVSG-----ESTGNCEDELPKGTREA---NGKKPSQQKFLERAPV-1359	
Dd	1882	PARYOFHIGSLERVQOLAIYPRESDSYEAVVPINFLCSRILYMSKEP---MKIYEITYRT	1938
Oy	1360	--NKSCDOIUL---KNTPDKKNNEHRESEKQO-----RTSTFOINGDNPK	1402
Dd	1939	TIONSSSTLTADVGRNYTYDHTNPNSKEVQLQMAQIARMHHTSLARSEPLENGSTDWGE	1968
Oy	1403	ITLKGECLKETISERVYSGNVERKVNNINKIIP---ENDIK-SUTKESAIRPINDEV	1457
Dd	1999	.PENPNSCVPP-----DONTEEBPOQADLLPMLTSPLEKFGLSTHGILLMLLVGV	2051
Oy	1458	IMEDFENENSSETKSHLSSDAAGNRDSELFP-----STKESDSTOTT	1504
Dd	2052	V----RLKQGELKDAlF--EDLPHELBDISMDLFIDDKTDLFAISEGSKDQMT	2105
Oy	1505	PSASCPESNSVNOVEDMEIETSRYKYVTSPIITSEESNLNDPIDENGPI-TKNENVN	1563
Dd	2106	S-----NQNONONOAGANVS-----ICDEBTRNSNTSLG-NGPASPNEVEDAM	2150
Oy	1564	GESKRKYITEVTWTSTVA-----TESKYIKYKGDQCVYVSFTNCACKSVITTT	1616
Dd	2151	LSAARN---SSQVOYMELTMPKLDGNSAMATIKRRKISK-----NLAEVGFILJS	2199
Oy	1617	TTVTLSTPSTGSGVDIIISVKEOSKYVVTTVDSLTGTLTGTLVMT--VSKEYSTRD-	1673
Dd	2200	SOORRKRMATAIVAGY-----SRQSISETSVGC VATTSGSVRSKSFYWSAAKRFRKSE	2255
Oy	1674	-----KVKLINFSPRPKTRSGTALPSYRKFVTKSTKSIFFVLPNDDLKTLARKGIR-	1755
Dd	2254	GREEAKARRIMQMGG-----VDDSITEPRIISGDONISTAQFSQVACD	2297
Oy	1736	--EYDFYNYNAKPLDLWPBPSPRPFGITWRIRLOTVYSLAGVSLMLMLMASLRMDM	1783
Dd	2298	RCOOCTRYNYDAFOR---HLPSCSPT-----MSENTESDVS-----	2330
Oy	1784	AAKVPGGGSGRRTSETSETITTEITTIKRRDVGYGIRFEYCIRKICIPIGVPEPKERTP	1843
Dd	2331	-----GGGMNNATQISABLSBNLOQLIANAGLN---CLO-----SATSF	2366
Oy	1844	PO-----RKGRSSALRPKR-----PETPKQTGVILETWABEEIEL	1881
Dd	2370	POVOSLSGLSGFGLGIQLOQLQPQSLGDBFSLSQDPRPADOT-----DDLOI	2418
Oy	1882	W--EIRAFAEV-----EKKAQAVEQAKRRELQEQPJYATISTPT	1923
Dd	2419	YANSLOSIALANLGSGFTLAQPTYAPAOPLIVASTNPBGTOOQFIQIPQTMATATTPTAT	2478
Oy	1924	SSTSTISPAOKVWVAPISSGYTTTKRKMYLT--TKGSPAITYTFQONKNFHOTFTATWK-	1960

Db 722 ADTTOJETVVDKSKKKVLK--TEKDSF-----ISQSETPPVEP 766
QY 777 NIPSEFHPWASHANWIKAYOMCSKRPFPALAL-ALIECAKPV--MLPTMREFLGH 833
Db 767 TKP-----ABSEAKIAEVNAKAKQKQVNDNLKREAVAKKIADKLEAE--- 814
QY 834 TRLHMTSIEREKKEVKKKKQKEEETMOQATWVYTPPVKHQVWKQGEERYATGYG 893
Db 815 ANIKTAIEVAAKKQKQKEDQKLETEVYVSKKAAEKLEKQAQI--KKAAL----- 865
QY 894 GMSWIKTHYRVPKLPCTNNVNRKSLFEGTKNNMDENK----- 933
Db 866 -----ADAVKQKQKELNENKLEAAKKSADKLEKEESAASKKVSSESVK 912
QY 934 DESDKRCSNPKKIKTEPDEKD-EVKGSDAKGADQ--NEMDISKITEKKQDYKELL 990
Db 913 GEKKTKRAGEKTYOVESEPTSKKTIDTKDVGATEPADETPKKIKKKTEKSSSISQKS 972
QY 991 DSDSDRCK---EEP-----MEYDDMKTESHV-----MCOES 1020
Db 973 ATDSEKYSKQKQEDPTKPAVSETQWTEADSKKQKQKEDKLEKLAELAKTKQEADEK 1032
QY 1021 SQVD-----VYNVSE-----GFHLRTSYKKKTKSSKLDGL-LERRIK-----QF 1058
Db 1033 SKIDAOEKIKKVEDDAARKEKELNDKLESEIATKKASADKLEEQQAQKAAVEVA 1092
QY 1059 TLEKORLEKIKLEGIGIKGTSTNSKNLSBPVITK---AKEGCSQSMRQEOSPNA 1115
Db 1093 AKQKQKEDQKLD-----TEASKKAQAACKLELEKQAOIKKKAAGADAVAKQKELDE 1144
QY 1116 NNDQPEDLIQGCQSQSDSVLRMSDPSTHTNKLVPKRDVLDDVSIRSPEKCPKQNSIEND 1175
Db 1145 KKK-----LEANKKSAAGKLIKIEESAASKQVTEQAKLDAQTKATAE--KQTKLEKD 1197
QY 1176 IEKVSDLASRQEPKSKTKGMDFTIDSKLASADIGTLCKNNKPLQIESDPTIVS- 1234
Db 1198 -EKSTESKSEKTVDEKPKK-----VLKKK-----TEKSDSISQ 1232
QY 1235 ---SSKSAHSSVPSKSTNDR--DATPLSRAMDFGKLGD-----SESNTL 1276
Db 1233 KSTSTKTVESAGPSESETKQVADARQKQETDEKQKLEELITAKKSADKSKLEESKL 1292
QY 1277 ENSSTVSIQDSSEEDMIVQNSNESISEQFRTREQDVEVLEPLKCELVSGESTGNCEDRL 1336
Db 1293 KKAAEVAEAKKQKQKEDQKLDTEAASKKAAEKLELKOSHUK----- 1336
QY 1337 PVKGTANKKKPSQKKLEER-----PVNKCSDQIKLANTDKN-----NENRESE 1383
Db 1337 --KAAEVDAVK--KQELEKQKQLESEAAKKAADAEKLEEQKKAAEIALIETQKEQ 1392
QY 1384 KKGQRTSTQINGKDNKPKIYLGKGC-LKEISPRVSGNVPEPKVNNIKTIIPENDIKSL 1442
Db 1393 KLAQBOBSRLEDEKKAQKQKLESETKSKQTEAPKESVDEKPKKVLKKTIEKSD-SSI 1451
QY 1443 TYKESAIRPPIN-GDVIMEDFN-----ERNSETRKSHLLSSSADGANYRDSLETL 1491
Db 1452 SQKSKAKSTVAEAETLESDFNLYEKKTQVQKQESPEDESTSATIKRDPAQ-----KTEEL 1506
QY 1492 PSTKESDSQOTTPPSACSESNVNOVEDMEIETSE--VKKTY--SSPTTSE-----EESN 1543
Db 1507 SKQDDSDDEKKTYY-----DGKPPKPEDESEATPKRKRKVKKTQKDSVASASLADAVSK 1559
QY 1544 LSNDFIDENGLPINKNENANGESKRTVTTEVTYMTSTVATESKTYIKVEKGDK----- 1597
Db 1560 LSDDVBEKKKKVKTKT-----EKSDSVISE--TSSVDTKIPES-VEIPLPEKAEQMLHNR 1613
QY 1598 ----QTVASSTENCAR-STVTTTTTTTTVKLSPTSGSVDIISVKEQSKTVVTTT----- 1647
Db 1614 FSTDSAVESEPKNAHKDTEKTTDDMMTRKSSA-----IFSDPQSISSKTSSEGRRR 1667
QY 1648 -----VYDSLTITGT-----LVTSMTYSKEVSTQDK-----VTL 1677
Db 1668 RRRGTGASKFASDITLALRGDNVEIEAEILLAEDDTYVWKVNGDADALNSRCHESHTFRT 1727

QY 1678 MKESRPKTKRSGTALBSYRKVFVTKTKSKSIFVLPNDDLKKAAGKIRE-----VPYEN 1731
Db 1728 LIIDEVEPTDSGMEITATGCTESHNTTLKVEELPVDFVYLPKRTSGKQGEVLTISVTIN 1787
QY 1732 YNAKPAOLDIMPYSPRPTGKITWRYLQTVKSLAGVSLMIRLLMASLMDMDMA--KVPPG 1790
Db 1788 HPIDISKVYW-LKDGRP-LEINKDYSIDYWG--CWSITLIR-----RAKEDSGKRYVQD 1839
QY 1791 GSGTRT-----ETSETEITTT-----EIKRRDNGPY-----GIREFYC 1824
Db 1840 GYDCSTHLSIQGKPVLKNSVETKPVITVDKDDQFSLIVAYDSNPESFMTVDGKLEFD 1899
QY 1825 IRKTIQPI-----GVPEPK-----ETPTPKRLBSALRKPKEPRTQGPILIE- 1871
Db 1900 GNSRIDVDDGLKTKRGVSKTIDAGEYKALKNEFEVQAKEDVKVNDTPSAPGDVSVK 1959
QY 1872 -----TWABEELELEIRAFARVEKEKA-----QAVEQAKR-----LEQOK 1911
Db 1960 AESDCLHIEWTAPTEEDNGAEVTSYV--IEKKESGRKKFKHVAIVNKKTSYVDDLEIET 2017
QY 1912 PVIYATSTSP-----TSSTTSTSPAQ--KVMAVPTSGSVTT----- 1947
Db 2018 PYIVRIAAVKKFQGTGEFIEFKPVQTSFPQVPTVEPPTIDVNTSYCSLSMPKPIEDGG 2077
QY 1948 -----GFKVLTTRKVGSPATVTFQOKNPNHQTFATWVKQ--QSN 1985
Db 2078 SPVYGIDVYKRENEGWMQKNGEELVFTSEFNRA--LSSGKEYEFKIEACNEAGLKN 2134
QY 1986 SGVVOVQKRVLAGTPS-----STGSOQFTSPQPTATVITPNTSG----- 2028
Db 2135 SNVYSKRLIYEGVLVPELIDMPKVLIDNDKVEVTWKSQGEKEFVQYQSDSISWASVD 2194
QY 2029 SGGTTSNSQ-----VITG-----POIRPG 2047
Db 2195 IGGPSESATSKCIIIDGLREGIPYEVRAARNQGTGFSEBPTIPVVYLADAPRVLKA 2254
QY 2048 MVIYRPTLOOSTL-----GKAIRT-----PVMQPGAPQOVMOQIING 2086
Db 2255 IKPVKIPKQGLRLBEHAAGHPAPEYIWKQKKEIIPTEBNEIYEGSMALIIHELGA 2314
QY 2087 QPVST-AVSAPNTVSTPQKSLTSATSTSNIOSSAOPRPOG--QVKLT----- 2135
Db 2315 EDVGLKVLVENVHGTAEBAEAVGISDVRAHFNSSSELTETIEEGHDILTCOVSEAV 2374
QY 2136 -----MAQITOL-----TQHGAGNGGLYV--- 2155
Db 2375 VMYKDKKRLVASDRVQFYAMARKRTLRIKSTADSDGVYKCEITDGRSRTGEVIVNQ 2434
QY 2156 -----IOGQGT-----TGOLOLIPQGVVLPGRGOOLMOAMPNGTV 2193
Db 2435 EPHIIVGPDALVKGDEGMVLFCEITSKPVYKVKPKKNQVEIETPQONKKAIMENDGRATL 2494
QY 2194 QRELTPLATFTATTAATTTTAVSTTAAGTGEOROSKLSPO-----MOYHODKTLPPAOS 2248
Db 2495 EIKNEFDKIDIGAVTASVSEKTSAPAKLVEVAPNLIIPTEIRDGTYVHAGNEFPA--- 2551
QY 2249 SVGPAAQOQOTQOPSAKRPQOTOPQSPAPQEVQTOPEVQTOYTVSSHVSEQPIHAOS 2308
Db 2552 -----VEFSGFPIPTIHLTNGTPLKAAIYAVEYDOSVVRM--KDVLTDSNG 2597
QY 2309 KPOVAASOPQSNVOGQSVVRVQSPQRTIRPSRPSOLSPQGOQOQVOTTSQPIPIPT 2368
Db 2598 TVRVIAES-PLQOCIKELIPLKIID-----KSAPODL-----QREVEYEDSVFLSMQ 2644
QY 2369 SLQIPSOQOPQOPQVQSSQOTLQNOYVSAPS--RPQLQIQOQOP----- 2417
Db 2645 --PLETNGAP-----LGVYIERKAVYDNNRMRPCGOYKPKPLRFGVADLPCN 2689
QY 2418 -----QVIAVPOLOO-----VQVLSQIOSVVAQIQAOQASVPQOIKQLPIQIQOSSA 2467
Db 2690 QVYGRITIAVNVGESEPCDITVLLTLESSEPVSS--ESSESEFVPKIATLRP--QVIVA 2744

QY 2468 VOTHQIOWVTVQAAVQOE-QLORYQOLR-----DQOQKKKQOQIEN----- 2509
DB 2745 VD---ETKVTILMEECPTSLYKVERKKVGDSDMLEIANTIRNRKFKDSTLSESEYVQ 2800
QY 2510 -----VNTPSKLLIYEIIOKOVVKNHNAIVIEHLKOKKSMTPAEREENOMIVCNV 2561
DB 2801 VTANGIHAVSSSEETNPVKILIVPGSEMPAS-----KTEKTDAAKSSSEKSS--AEET 2852
QY 2562 MKYILDKIDKEEQAARRKREESVEOKRSKOMATKLSALFLFKHEQOLRAELIKRRALLD 2621
DB 2853 V-----AEKQVDSQASESTTEAVEEKKTK-----KVYKKKVAEN 2887
QY 2622 KDOI--EYQELKRDILIKKENDLMQOLAQATAVAAPCPVPVLPVLPAPPPPPPPPPPG 2679
DB 2888 KGEETLOVYKELKKGKAVKQVQDSRGRSLQASS----- 2922
QY 2680 VOHTGLSTPLTVASOKRRKREESKSSKSKKKMIS-----TTSEKTKDT 2727
DB 2923 -----DNESVTTTSEKRESEAELEKNSKSAKKSTSDLEADKAETEKSETGKET 2973

RESULT 9
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8221122
A:Accession: B48666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SOCH>
A:Cross-references: EMBL:X65551
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KiH>

Query Match 3.0%; Score 446; DB 2: Length 2897;
Best Local Similarity 18.7%; Pred. No. 9.7e-09;
Matches 559; Conservative 398; Mismatches 1058; Indels 982; Gaps 140;
QY 378 CKILEEMREIHRMDITEDLTN---KARGSNKSFLAANEELIESIRAKGDDIDNYS 433
DB 49 CKI-----EIHGEAILHNESSTNPQVNGS-----VIDEPYRLKHGDV--ITI 90
QY 434 PETEKDKNET-ENDSKDAE---KNREEFEDOSLEKSDDKTPDDPQOKS-EETPEVG 488
DB 91 IDRSFRYENESLONRKSTTEPRKIREQEPARVRSFSSPDDEKQADQSKAYSKITEGK 150
QY 489 DKGN--VSANLGDNTNATSEETSPSEGRSPVGLSETPPDSNMAEKKVASLEQDVE 546
DB 151 VSGNDELDEENPPVTPPLKRG--APYRKSLV---MHTP---VLKRIKEQPPSGK 201
QY 547 EPNKCESSNTSATTTSTIOPLENSSSELNSQSESAKADDPENGERESHETPVSIQ 606
DB 202 QESGEIHVEKQASLVSPAPSPRKTTPVASDQRRRCKTA-----PAS--- 246
QY 607 EIVGDFTSEKSTGELSESPGAGSGSTRITIRLNPDS-----KLSQLKSQVAAA 660
DB 247 -----SSKSQTEVPK-----RGGERVATCLOKRVSISRSOHDILQMICKRRSGA 291
QY 661 HEANKLFEGEVLVNVSGEISRLTKK---EVTMKG--NINNYFKLGSGKRV--Y 712
DB 292 SEAN-----LTVAKWADVVKLGAKOTQTKVHKGPQSRMNRORRPAATPKRPVGEV 343
QY 713 HNOYSTN-----SFALNKHQREDHDKRR--HLAKHFLCLPAGFEK-----WNGS 755
DB 344 HSQFSTGANSPTIILICKAHTKHYHAPRPYVNLNFIISNCKMFKEDLSIAEMFKTP 403

QY 756 VHGSKVLTISTRLTITQLENNIPSSF-----FHPNMAASHRANWIKAVOMSK-PR 805
DB 404 VKEQDPLT-STCHIAISNENELGKQFOGTGCEBLLPTSESFGVNEFSAONAKOPS 462
QY 806 EFALMALALECAVKKPVMLPIWREFLGH-----RLHRTSIRE----- 845
DB 463 D-----KCSASP---PLRQCIRENGVNAKTPRNTYKMTSLETKTSIDETERSKTV 510
QY 846 -----EKEKYK--KEKKOEELTM-----QATWVYTPPVKHQVKKOG 884
DB 511 STVNSGRSTEFRRNIQKLVESKSEETNEIYECILKROKQATLLQ-----RRG 561
QY 885 EERYPTGYGWSWISKTHYRFRVPLPGNTVNYKSLSEGTNNMDENDESCK-----R 939
DB 562 EKKEI-----ERPFET-----YKENILAKEN--DEKKAMKRSRTWQ 597
QY 940 KCS--RSPKIKITPDSE--KDEVKG-----SDAKGADQNMDSKTEKKDOVKEL 989
DB 598 KCAPNSDLTDLKSLPDTLEMKDTPARQONLLOTODHAKAPKSEKKITTKM----- 646
QY 990 LDDSDKPKCE-EPMEDDDMKTESHVNCQESSQVDVNVSE-----GFHLKTS----- 1037
DB 647 -----PCQSLQP---EPINTPTHTKQQLKASLGKVGKVELLAVGFTTSGETHT 695
QY 1038 -----YKKTKR-----SKLDGLERIKQOTLEKQKLEKIKLEGGIKIG 1079
DB 696 HREPAQDGKSIRTFKESPQIIDLPAARYGM--KKWPRPKEEAOSLEDL---AGKKELF 750
QY 1080 KRTSNSSKNLSSPVYTKKCEGQSDSMRQEQSPANNDQPE-DLQCGSQSDSVLRMS 1138
DB 751 QPFGSEEMDEKT---TKACKSPPEPVPTPISTKQMPKRLKADVEEFPLAKRL 807
QY 1139 DPSHTTNKLYPKDRVLDVSI-----RSPETKCPKONSLEN--- 1174
DB 808 TPSAGAKAMLTTPKAGDEKDIKAPMGTPIVQKDLIAGTLPGSRQLOTPREKQALEDLAG 867
QY 1175 -----DIEKVSDDLARGO-----EPTSKTYGNDEFIDDSKLASADIG 1214
DB 868 FKELQTPGHTEIELVAAGCTKTRICPDSPOSDVDPTSTYKOR-----PKRSIRRADVG 921
QY 1215 TLI-CNNKPELLOESDPTIVSSKSAIHSVPKSTNDRD---ATPLSRAMDFESKILG 1268
DB 922 ELLACNLMIP-----SAGKAMHTPKPSVGEHEKTIIFVGPVOKL----- 961
QY 1269 DSESNSTLSSDPTVSIQDSSSEDMIVONSNESISEOFTREQDVEVLEPLKCELVGES 1328
DB 962 -----DLTENLNGSKRRQPTPKREQALEDLIGFKELPOTPGHTBEA-----VAAKGT 1009
QY 1329 TG-NCEDRLPVKGTBANGKRPQOAKLEERPYNKCSDOIKLNTYDKKNENRESEKGO 1387
DB 1010 TKMPCSSPPESADPTSTRQPKTPLEKRDYQKELSAK-----KLQTSGE 1057
QY 1388 RTSTPOINKOKMKPKIYLKGECLKELSESRYVSGNV-EKVYNNIKITIPENDIKSLTYE 1446
DB 1058 THTHTKVPGEDEKSIINAFETAKQKIDPAVATGSRHHPKTR--EQAQLELDAGN--KE 1113
QY 1447 SAIRPFINDVIMEDPFENENSETKSHLLSSDADAGNYSLETLPTSTKESDSTQTTPS 1506
DB 1114 LFGTP-----VCTDKPTTHEKTKIKACSQPD-----PYDPT 1146
QY 1507 ASCPESNSVNOVEDMEIETSEVKKVYSS-----PITSEESNLNDPIDENGLPIN 1557
DB 1147 SSKPQSKRSLRKVDVEEDEFALLKRTTPSGAKAMHTPKPAVSEKKNIYA--FM---GTPVO 1201
QY 1558 K---NENNVGSEKRTKTVITEVITMTSTVA-----TESKTVIKVEKGDQOTVYS 1602
DB 1202 KDLTENLNGSKRRRLQTPREKAQALEDLAGEKELFOTRGHTESMT---NDKTAIVACKS 1258
QY 1603 STENCAKSTVTTTITVTKLSTPSTGSGVDIIVSEKQSTVTTTITVDSLTITGGTLVYS 1662
DB 1259 SQPDLDDKPNPASKRRLKLSLG--KYGVKEELLAAGKLVQTSSETLHTHTHPFGDGSMA 1316
QY 1663 MTVSKREYSTROKVKLMEKPSRPKRTSGTA-----LPSYRK--FVTKSTYKS 1706


```
Db 1317 FMESPFOILDSASLIGSKROLTRPKGSEVEDLAGFIELOTPSHTESMTNEKTV 1376
Qy 1707 IYVLPBDL-----KTLARKGIREYFENYNNKRALDIMPYSPRPPTGITWR 1755
Db 1377 STRASOPDLVDTPTSSKPOPKRSKRKADTEE-----ELAFKQTPSAG---- 1420
Qy 1756 YLQIVKSLAGVSLMR-ILMASLFRMDMAKVPFGSGSTRTETSELETTEITTEILRRDV 1814
Db 1421 KAMHHPKPAVGEKDIINTFLGTPVOKLDOPNLPGSNRRLOTREKKAOLEELT---- 1474
Qy 1815 GPYGR-----FEYCIRKIC--PIGVN-EPKKEPTPPQKGLASS----- 1852
Db 1475 ---GFELEFQTPCTDNPTADEKTYKKILCKSPQSPADPTPTMTKOPKRSKLKADVEEF 1531
Qy 1853 -ALRPRP-----ETPKOTGVIIETWVAEELELELRAF---AEVE----- 1892
Db 1532 LAFRLKTPSAGAMHPKA-----VGEER---DINTFVTPVEKLDLGNLPGSK 1579
Qy 1893 -----KEKAQAVEOQAKRLEQOKPTVIATSTSTSTSTISPAQKVMAPISGVY 1946
Db 1580 RRPQTPKEKAKLEDLAGKELEFQTPGHTESMTDKITIEVCSKQPPDPVKT----- 1633
Qy 1947 TGTKNVLTTRKVSPPATVTFQONKNPHQTFATWVKQOSNGVVOYQKVL--GIIPSSTG 2004
Db 1634 TSSKORLKISLKG-----VGYKEEVLPGKLTQTSG 1664
Qy 2005 TSOQFTSQPRTATVYTRPNTSGSGT-----TNSQVI-----TGQIRGMVITRT 2053
Db 1665 KTTQT-----HRETAGDOKSIKARESKAKOMLDBANTGTGMRP-----RT 1706
Qy 2054 P-----LOOSTLGKALITTPYVWOPGAPQVMTQITRQOPSTAVSANTVSTFGQKS 2107
Db 1707 PKEEASLEDLAGKELEFQTPGHTESTDDKTKT-----ACKSPPPS 1751
Qy 2108 LFSATSTSIQSSASOPRRPQGOYKVL--MAQLTQITGHCNGGLVIVIOGQCGTQO 2165
Db 1752 MDTPTST-----RRRPKTPLGKRDIVELSLAKOLTO-----TTH 1786
Qy 2166 LQLIP-----QGYVYLPBGQOLM-----QAMMNGTVQ-----RFLP--TPL 2201
Db 1787 TPKVPEDEKGINVRETAOKKLDPAASVTSKROPRTPKGAQPLEDLAGLKELEFQTV 1846
Qy 2202 ATTATYASTTTTSTTA-----AGTGE--QROSKLS--POMOVHODKTLPPAOSSSVGA 2253
Db 1847 CTDKPTHEKTKIACRSHPQDPVGPPTIFKQSKRSKLKADVEESLALRKRTPSVGA 1906
Qy 2254 KAQOTA-----QBSARQOTOPQSPAPQEVQTOPEVOTOTTSS 2294
Db 1907 MDTPRKAGDEKDKAFMGTPVQKLDLPGNLPGSKRMPTQPRE-----KAQALEDLAG 1959
Qy 2295 -----HVPESEAPTHAOSKPOVAASOPQSNVQSGSPRVQSPQTRIRPS----- 2341
Db 1960 KELEFQTPETDKPT--TDEKTKIACKS--PQ-----PPVPTPASTKORPPRNLRKADV 2010
Qy 2342 -----TPS-----QLSPGOQOVQTTQSPPIPIQPHTSI--QIP--SOGQPOSQP- 2382
Db 2011 EEEFLALRKRTPSAGKAMDTPPKPAVSDEKINIMTVEYVOKLDLGNLPGSKRQTPRE 2070
Qy 2383 -----QVQSTOTLSSGOTLNOVSVPSPRQLOIQOPQYIAPV----- 2423
Db 2071 KAEALEDLVGELEFQTPGHTESMTDDKITIEVCSKSPQSEFKTSRSKQRLKIPLVKY 2130
Qy 2424 QLOQOVQVLSOI--QSOVAAQIOAQOQSVPOQIKL--QLPIDIQSSAVQTHQIQVNVY 2479
Db 2131 DKKEEPLAVSKLRTSGEYTOHTERTGDSKSIKAFKESFQKILDBAA-----SVTG 2182
Qy 2480 QAASVOEQLQRYOOLRDOQOKR-----QOOIEINVT-----PSKLLIKVEIIO 2524
Db 2183 SRROLRTREKRALEDLVGELEFSAPGHTESMTIDKNTIPCKSPPELDTATSTK 2242
Qy 2525 -----KOVVKNHNAVIEHLQOKKSMT-----PAEREENOMIVCNOVMKTYIIDKIDK 2571
```

```
Db 2243 RCPKTPRKEVEKEELSAVERLQOTSGQSTHTHKEPASGDEGIKVL----- 2287
Qy 2572 EEKQAKKRRKREBSVQKRSKQNA-----TKLSALFLFKHEOLRA----- 2611
Db 2288 --KQRAKKKPNVEEPPSRRRPAPKEKAQPLEDLAGFIELSTSHTOESLTAAGATKI 2345
Qy 2612 -----ETLKRKALLDKDLQIEVDEELKRLDKIKKEKDMLOLAQ----- 2649
Db 2346 PCSPPELEVVDITASTKRLHRTFYQK-----VOVKEPSAVKFTQSGETTDADKEPAGE 2400
Qy 2650 -----ANVAPRCPPYVPLPAPAPPPSP-----PPQVQHGLSTPPL 2691
Db 2401 DKGIKAKESAQTPAPASVYSGRRRRPAPRASAQLEDLAGFKDPAAGHTEESMTDK 2460
Qy 2692 PVASQKRKEEEDSSKSKKMKMISTSKETRKQDKLVCICQTPDESKFYIGCDRCQN 2751
Db 2461 TTKIPCKSSPELEDYATSSKRRRRTQAQVEKEE--LAVGK----- 2501
Qy 2752 WYHRCVGILOSEALELIDEVYVPOCOSTEDAMTVLPLTEKDYEGLRVLRSLQAHKMAW 2811
Db 2502 -----LTQTSGE-----TTHDKPEYGEKGTGA----- 2525
Qy 2812 PLEPYDPN-DAPDYGYATKEP-----DLATMEERQVR--RYEKLTFVAD 2856
Db 2526 -FKQPAKRVVDAEDYVSGRRRQAPRPAKEKAQPLEDLASFQELQTPGHTELANGAD 2581
RESULT 10
F90073
Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-U, T.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratazu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447
Query Match 2.9%; Score 440; DB 2; Length 2271;
Best Local Similarity 18.3%; Pred. No. 1.2e-08;
Matches 442; Conservative 332; Mismatches 1000; Indels 642; Gaps 84;
Qy 411 AANEEILEISAKKDDINNVKSPETEKKDKNTEKDSMDAKENRREFDQSLKNSDDK 470
Db 90 AASDAPLITSELTQSETVGNQNS--TTLEASTSTADTSVTKN-----SSSVQTSNSDT 141
Qy 471 TPDDDEQKSEEPREYVGDKGNSVANLGDNTTNATSE--TSPEGRSPVGCLESTP 526
Db 142 VSE-----KSEKVT-----STTNSISNOQKLTISSESTSS--KNTT 177
Qy 527 DSSNNAKKVASELPDQVDEEPKTCESSTNTSATTTSIQPNLNSNSSLSLNS----- 579
Db 178 SSSDTKSVASTSTEQPINTSTNGSTASNNSTQSTTPSSVNLNKTSTSTPAVKLRTP 237
Qy 580 -----SQSESAKADDPENGRESHTPVYSIOBELY-----GDFISEKSTG--ELS 622
Db 238 SRLAMSTFASAA-----TTTAAVTANTITVKNQNLKQYFTSGNATVYDOSTGLVITL 288
Qy 623 ESPGAGKA--SGSTRITRLRNPDSKLSQKSOQVAAAHANKLFEKGG--EVLVNS 678
Db 289 QDAYSQKGAITLGTRI-----DSN-----KSFHPSGKNVNLGNKKYEGHNGGDGIGFANS 337
Qy 679 QGEISRLSTKKEVINKNINNYFKLGQEGKRYRVYHNQYSTNSFA----- 722
```


cell proliferation antigen Ki-67, long form - human
C.Species: Homo sapiens (man)
C.Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C.Accession: A48666
R.Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A.Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq-
ins.
A.Reference number: A48666; MUID:94043435; PMID:8227122
A.Accession: A48666
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-3256 <SC>
A.Cross-References: EMBL:655550; NID:9415818; PIDN:CAA46519.1; PID:9415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F.29-91/Domain: kinase interaction domain homology <Kih>

Query Match 2.9%; Score 439.5; DB 2; Length 3256;
Best Local Similarity 18.3%; Pred. No. 2e-08;
Matches 579; Conservative 432; Mismatches 1103; Indels 1049; Gaps 148;

```
QY 300 VPGVTCVAVICQKNKRYIRHPIGYDRSKRYFLNKRLLIEDTENENKKTWY----- 354
DB 221 VP--TTCCLDNKSKNE-----SPEWKLYESVKKELDVKSOKENVLQYCRKSG 265
QY 355 ---YSTQVLAELIDCLDKDYWEALCKILEEMREIHRHMDITPDLNFKARSGNSFL 410
DB 266 LQTDVATEKESADGLO-----GETOLL-----VSRKSRKSGSGSHANA 304
QY 411 AAAN-EELESTRAKKGDIDNYSKPEET-----EKQKNET-----ENDSKDAEKNR 455
DB 305 EPASPEQELDONKGRKRDVESVOTPSKAVGASEPLTEPAKMKTPVQYOOQONSPQKHKK 364
QY 456 EER-----EDOSLEKD-----SDCKTPDDDEEGCKSEEPVEVQKGSVS--ANLGNNTN 504
DB 365 DLYTTTRRSVNLGKSEGRKAGDKITLPPKRLSTRNTPAKVEDADATKPEVLSKTRG 424
QY 505 A--TSEETSPSE--GRSPVGLSET-----PDSSNAEK-----KVASELPQDVEP 546
DB 425 SIPTDVEVLPTETELINEPFLTLMLQVERKIQKSLKPEKICITAGOMCSGLPLSSV 484
QY 547 EPRKTCESNTS-----ATTSTIOPNL--ENSSNSSELNSQSASAAA----- 588
DB 485 DINNFQDSINSEGLPLKRRRVSEFGHLRPELFEDENLPPNPLKREAPTKRKSILVMPH 544
QY 589 -----DDPE-NGERES-----HPVPSIOEETVG-----DPTSKSGSELSSEPG 626
DB 545 PVLKTIIRKQPOPQSGKQSGSEIHVYKQSLVISPAPSPKTPVASDOORRSCKTATA 604
QY 627 AGKGAS-----GSTRILITRLRNPDS-----KLSQLKSQOYAAAAHANKLFKEGEVL 674
DB 605 SSSKSGQTEVPKRGGERVATCLOKRVASIRSQHDILQIMICSRRSGASEAN-----LI 656
QY 675 VYNSGGEISRLSTKK--EVIMKG--NINNTFKLGQBEKRYV--YHNOYSTN-----S 720
DB 657 VAKSMADVVKLAKQOTQTVIKHGPQSRNNKQRPAPRPATPKKPVGEVHSGFSGHANSPT 716
QY 721 FALNKHQHRHEDDKRR--HLAHRFCLTPAGEFK-----WNGSVHSGKVLITLTL 769
DB 717 IILGAHTEKVAVPAPRYVNLNFIISNOKMDFKEDLSGIAEMFKTPVRKQPOLT-STCHI 775
QY 770 TITTOLENNITPSSF-----FHPNMAASHRANWIKAVOMCSK--PREFALALALIECAV 819
DB 776 AINSNENLQKQFOGDSGEBPLPTSESGFNGVFEFSAONAKOPS-----KCSAS 827
QY 820 PYVMDPIWREPLIGHT-----RLHRMSTIERE-----EKKK 849
DB 828 P-----PLRQOCIRENGNNAKTPRNTYKMTSLTKTSDTETTESKTVSVINRSGRSTEF 883
QY 850 VKR--KEKKOEDEETM-----QOATWVYTFPVKQVVKOKGEEVRYRVYGGWSMI 898
DB 884 ICKLPEVESSEETNTEIYVCILKRGOKATLLQO-----RRGEMKET----- 925
```

```
QY 899 SKTHYVRFVPKLEGNNTNVYKRSLEGGKNNMDENMESDK-----RKCS--RSPKKIKTE 951
DB 926 -----ERPET---YKENIEUKEN--DEKMAKMRSTRWGOKCAMPMSDITLTKSL 970
QY 952 PDSE--KDEYVG-----SDPAKGAQONEMDISKITEKKQDVQKELLDSDPKCKE-EP 1002
DB 971 PDEELMKVDARQONLQOTQDHAKAPKSEKCIITKM-----PCQSLQD 1012
QY 1003 MEYDDMKTESHVNCQESSQVDVNVSE-----GPHLRTS-----Y 1038
DB 1013 ---EPINPTHTKQOLKASLCKVGYKVEELAVGKFTPTSGETTHHBPAGDGKSIKTF 1068
QY 1039 KKKTK-----SSKLDLLEIRIKQFLPEKQREKIKLEGIGIKGICTSTNSKNLSSEP 1093
DB 1069 KESPKQILDPAAKRVGM--KKMPRTKEAQSLIEDL--AGFEELFOTGPSEESMTDK 1123
QY 1094 VITKAKEGCOSDSMRQEOSPNANNDOPE-DLIOGCSQSSSVLRMSDPHTTKLYPKDR 1152
DB 1124 T---TKIACKSPPEESVDIPTSTKQWPKRSLKADVEEFIALRKLTPSAGKAMLLPKPA 1180
QY 1153 VLDDVSI-----RSPETKCPKONSTEN-----DIEKVSD 1183
DB 1181 GGDEKDIKAFMGPVQKLDLAGTLPGSKROLQTPKEKAQALIEDLAGFKELFQTPGHTBEL 1240
QY 1184 ASRGO-----EPTKSKTGNDFFIDSKLASADDIGTL--CKNKKPLIQE 1227
DB 1241 VAAGKTTKIPQDSPOSDPVDPTSTKQK-----PKRSTRKADVEBELACRLMP----- 1290
QY 1228 ESDTIVSSSKSALHSSVPKSTNDRD-----APPLSRAMPFEGKLGDSSESTLSENSDT 1282
DB 1291 -----SAGKAMHTPKPVGEEKDIIIFVGPVQKL-----DLTELTGS 1329
QY 1283 VSIQDSSEEDMIVQNSNESISBQFTRQEDVLEFLKELIYSGESTG--NCDRLPVKCT 1341
DB 1330 KRRPQPKKEAQALIEDLTFGFKELFQTPGHTEA-----VAAGKTYKMPCCSSPESAD 1382
QY 1342 EANGKRPQOQKLEERPVRVVKCSQDQIKLNTTKKNNNESESEKCGORTSTQINGDNKP 1401
DB 1383 TPTSTRQPKTLEKRDVQKELSAK-----KLTQSGTTHTDVAPGGEKS 1430
QY 1402 KTYLGECKLEISESRVSGNV--EPKVVNINKIIPENDIKSLTVESAIRPINDVIME 1460
DB 1431 INAFRETAKQKLDPAVASYGSKRHPTK--EKAQPLELAGW--KELFQTP-----VCT 1480
QY 1461 DENERSSETKSHLSSSABEGNYRDSLETLPTSKESDQTQTTPSASQPEESNVQYED 1520
DB 1481 DKPTHEKTTKTAQRSQP-----PVDTPSSKPOSKSLRYVD 1519
QY 1521 MEIETSEVKKVTS-----PITSEESNLSDNFIDENGLPIPK--NENVNGESKR 1568
DB 1520 VEEFFALKRTPPSAGKAMHTPKPAVSGEKNIYA--FM--GTPQKLDLENLGSKR 1574
QY 1569 KTVIIEVTITMTIVA-----TESKTVIKVECKQOTVVSVTECNKAKSTVTTTT 1616
DB 1575 IQTPKEKAQALIEDLAGFKELFQTRGHTESMT--NDKTAKAAKSSQPDLLKNPASSNR 1631
QY 1617 TTVTKLSTPSTGASVDIIVKQESKTVVTTVTDSLTITGGTLVSMVSKYSTRDYK 1676
DB 1632 RLKTSIG--KVGKBEILLAVGLTQTSGETTYHTHEPTQDGSKMAFMSPOIILDSN 1689
QY 1677 LMKFSRPPKTRSGTA-----LPSYK--FVTKSTKSIFVLPNDL----- 1715
DB 1690 LTGSKROLTTPKGSKEVPELDLAGFTELFPQPSHTKESMTNEXTKTVYSRASQDPLVDPT 1749
QY 1716 -----KTLARGGIREVYENYNAKPAIDIMPYSPRPTETITMYRKLQTYKSLAGSL 1769
DB 1750 SSKPQPKSLRKADTEE-----EFLAFKQTPSAG--KAMHTPKPAVGEK 1793
QY 1770 MLR-LIMASLRMDMAAKVPPGGSGSTRPTSETETITTEIIRROVGPYGR----- 1820
DB 1794 DINTLGLTPVQKLDQPGNLP--GSNRLQTRKEKAQALBELT-----GFELEFQTPCT 1844
```



```

QY 1821 -----FEYCIKRIIC--PIGVP-ETPKETPTQKGLRSS-----AARPKP----- 1859
DB 1845 DNPTRDEKTKKILKSPQSDPADPTNTKQPKRSKLKADVEEELARLKLTPSAGKAM 1904
QY 1860 ETPKQGTGPIIETWAAEELMEIRAF-----AEREE-----KEKAQAVE 1900
DB 1905 KTKPKAA-----VGEER-----DINTFVGTEKLDLGNLPGSKRRPQTKREKAKALE 1952
QY 1901 QOAKRRLDQKPTVIATSTSTSTSTSTSTISPAQKVMAPISGSVYTGKRWLTTKVGSP 1960
DB 1953 DLAGEKLEFQRTGHTKEESMTDKITEVSCSKSPQDPVKTP-----TSSKQKLKISLKG- 2005
QY 1961 AIVTQOQKNFQRTATWVKQGSNSGVYVOQKVL--GIIPSSGTSGOQTTPSHQPRTA 2018
DB 2006 -----VGKKEEVLTPVKELTQTSKTKTQT----- 2028
QY 2019 TVTIRPNTSGSGT-----TSNSOVI-----TGPOIRGGMVIRRP-----LQOSTIG 2061
DB 2029 ---HRETRAGDGSIKAKRESKQMLDPANVTGEMERP-----RTPKEEAOSLEDLAGF 2079
QY 2062 KAIIRTPVWVGAPQOVWTOIIRGQPVSTAASAPNTVSSPTGOKSLNATSTNSIOSA 2121
DB 2080 KILFQTPHTEESTTDDKTKTI-----ACKSPPEEMDTPTST----- 2117
QY 2122 SOPRRPQOGVKLT--MAOLITLQGHGNGGLTVYIGOGQTTGOLLIP-----QGVTV 2175
DB 2118 RRRPKTPLEKRDIVEELSAIKOLTO-----THTTDKPGDEDKGINV 2159
QY 2176 LFGPQOQLM-----QAMPRNGIVQ-----RFLF-TPLATATATASTTTTV 2215
DB 2160 FRETAKQKLDPAASVTSKGRQRTPRGKAQPLEDLAGLKLELQTPVCTKPTTHKTKTKI 2219
QY 2216 STTA-----AGTGE--QROSKLS-POMVQODKTLRPAOSSSVGPAKAQOTA----- 2260
DB 2220 ACRSPQRPDVGPRTIFPKPQSKRSKLKADVEESLALRKRTPEVGKMDPRKAGBDEKDM 2279
QY 2261 -----QPSARPOQPOPOSAPQEVOTQPEVOTQTVSS-----HVPSEAQPT 2303
DB 2280 KAFKGTPTYQKLDLPGLPLPGSKRPQTPKE-----KAQALEDLAGFKELFQTPGDKPT 2332
QY 2304 HMOSSKPOVAASQOSQNSVQSPVAVQSPSQTRIRPS-----TPS- 2344
DB 2333 -TDEKTTKACKS-PQ-----PPVDTPASTKQPKRKNLKRADVEEELALRKRTPSA 2383
QY 2345 -----QLSPGQOSQVOTTSQPIPIQPHSL-QIP-SOGQPOSQ----- 2382
DB 2384 GKAMDTPKRAVDEKININFTETPYQKLDLGNLPGSKRQRPQTPKKALEDLVGFKE 2443
QY 2383 -QVQSTQTLSSGQTLNOVSVSSPKRQLOIQOPQVIAPV--QLQOQVVLISQI-- 2435
DB 2444 FQTPGHTESMTDDKITEVSCSKPOPESEFKTSRSSKQRLKIPLVKDKMEELAVASKLNR 2503
QY 2436 QGVVAQIOAQOSQVPOQIKL-QLPITQIOQSSAVQTHOIONVYVQAASVQEQLOVQO 2493
DB 2504 TSGETQTQTEPTGDSKSIKAKRESKQILDPA-----SVGSRQLRTRKEKAKA 2555
QY 2494 LRDQOQKK-----QOQIEINVT-----PSKLLIKVEIID-----KOVVMH 2531
DB 2556 LEDLVDFKLEFAPGHTESMTIDKNTKIPCKSPPELDTATSTKRCRTPRKVKKEE 2615
QY 2532 NAVIEHLKQKMT-----PAERENQMIYONQVMKYLILDKIDKEQKAAKKRRRES 2585
DB 2616 LSAVERLTQTSQSTHTHKEPASGDEGIKVL-----KQRAKKKNPNPVE 2658
QY 2586 VQQKSKQNA-----TKSLALLFKHKQQLA-----EILKKRA 2618
DB 2659 EEPSPRRPAPREKAQPLEDLAGFTELSTSGHTQSLTAGKATKIPCSPPLEVVDTTA 2718
QY 2619 LLDKDIQIWELELKLKIKKEKDKMLQALO-----ATAV 2653
DB 2719 SKRHLRTVQK-----VQVKEEPSAVKFTQTSGETTDDADKEPAGDKIKALKEKAKOT 2773
QY 2654 AAPCPVLTFLPAPRAPPSPP-----PPGVQHTGLLSTPLPVAASQKRRKEEKD 2705

```

```

DB 2774 PAPAASVTSRRRPRAPRESAQAIEDLAGFQDPACHTESMTDDKTKTIKSPSELED 2833
QY 2706 SSSSKKKMMISTTSKTKETKDKLYCICTPYDESKFYIGCDRCQNMVHRCVGILOSEA 2765
DB 2834 TATSSKRRPRPTAQKVEKAE--LLAVCK-----LTQTSG 2866
QY 2766 ELIDEVYCPQOSTEDAMTVLPTLEKDYGLKRVLRSLQAHKMAPLEPYDPN-DAPD 2824
DB 2867 E-----TTHIDKEPVGEGKTKA-----FKQPAKRNVDAD 2897
QY 2825 YGVVKEPM-----DLATMEERVQR-RYKELTFEVD 2856
DB 2898 VIGSRQRPAPREKAQPLEDLASFQELSQTPGHTEELANGAD 2940

RESULT 12
S37431
N:Ankyrin 2, neuronal long splice form - human
M:Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythro
N:Contains: ankyrin 2, short form
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence-revision 06-Jan-1995 #text-change 13-Aug-1999
C:Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569
R:Chan, W.
submitted to the EMBL Data Library, September 1993
A:Reference number: S37431
A:Accession: S37431
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3924 <CH>
A:CROSS-references: EMBL:226634; NID:9406287; PIDN:AA62828.1; PID:9406288
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A:title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal
A:Reference number: A39643; MUID:91302466; PMID:1830053
A:Accession: A39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2077 <OT>
A:CROSS-references: GB:X56957
A:Accession: B39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1443, 3585-3924 <OT>
A:CROSS-references: EMBL:X56958
R:Itse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Wa
Genomics 10, 858-866, 1991
A:title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A:Reference number: A40334; MUID:92009921; PMID:1833308
A:Accession: A40334
A:Molecule type: DNA
A:Residues: 463-474, 'BE', 477-495 <TSE>
A:CROSS-references: GB:M37123; NID:9178647; PIDN:AA62828.1; PID:9178648
R:Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A:title: 440-kD ankyrinB: structure of the major developmentally regulated domain and
A:Reference number: A49462; MUID:94075409; PMID:8253844
A:Accession: A49462
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3924 <RES>
A:CROSS-references: EMBL:226634; NID:9406287; PIDN:CAA81387.1; PID:9406288
C:Genetics:
A:Gene: GDB:ANK2
A:CROSS-references: GDB:127607; OMIM:106410
A:Map position: 4q25-4q27
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing
F:2-3924/Product: ankyrin 2, long form #status predicted <MA>
F:2-1443, 3585-3924/Product: ankyrin 2, short form #status predicted <MA2>
F:63-95/Domain: ankyrin repeat homology <AN01>
F:96-128/Domain: ankyrin repeat homology <AN02>

```



```

Db 2604 --SKRLRTHKGMSTLLKLLGSKETQ--1SDHSEKLAHDTSILKSTQOQR--DS 2655
Oy 2552 NQRMIVGNQYWK-----YLLDKID-----KEEQQAQRK--REESVEQKRS--- 2591
Db 2656 VKPLRTRRVRLASKEVPELVLDTRDHTLOSXNPLSPKRRKSGARPSIYRTRLSL 2715
Oy 2592 --KQNAKRLSALLFKHKEQJRAEILKRLALDLOI-----EVOEELK-----RDLKI 2638
Db 2716 APKQEAETDEKRVPEKKAASKRSVSPPEPKMHLLYSNKLSEVBOVSIVYMKTEMEA 2775
Oy 2639 KKEKDLMLQAQATAVAPACBPVT-----VLPAAPAPPSPPPPGVONTGL--LSTPLP 2692
Db 2776 KREN-----PYTPDQNSRYRKKTIVKQGRPFQSAENWGIKKNEKTMK 2819
Oy 2693 VASQRRKREEEKDSSSKSKKKMIS--TTSKEKDKTALCYCTCKPTPYDESK 2741
Db 2820 TASQETELQNDQQAACKSTRGVSSKRCJCLNSRGTEETPOCEAEKERTSK 2870

```


QY 1770 MRLMALRMDMAKVPBGSGSTRT-----ETSETETTTETIIRRDVGPYGINREYCI 1825
 Db 1542 -----NEROQNTGLKNTPSEGOQNTGLKNTPSEGOQNT----- 1574
 QY 1826 RKIIPIGVPEPKETPIRQKRLSSA-----LRKRPETPKQGPVILIEIHWAE 1877
 Db 1575 -----GLKNTPNE--ROQNTGLKNTASKGOQNTGLKNAPNEROQNTG--LKNTPSEG 1622
 QY 1878 ELEMEIRAFARVEKE--KAAVEOQAKRL-----EOKPVIATSTSTSTST 1929
 Db 1623 Q-QNTGLKNSASKGOQNTGLKNTPSEGOQNDLKNAPNEROQNTGLKNTPSEGOQNTGLK 1681
 QY 1930 ISPAQ-----KVMAPISGSVTGTGMVLTGVSPATVTFQONKNEHOTPATWVKQGS 1984
 Db 1682 NTPSEGOQNTGLKNTPSEGOQNTGLKNTPNER-----QONTGLKNT--PSEGOQ 1728
 QY 1985 NSGVYQV---QOKVIGIIPSSGTSGOQFTSFQPKTATVITRPNTSGSGTSSNSOVLITG 2041
 Db 1729 NTGLKNTNEROQNTGLKNAANKGOQNTGLKNTPNEG---QONT-GLKNTPSEGOQNTG 1783
 QY 2042 POIRPGMTVIRTPLQOSTLGKAIIRTPVMWOPGAPQ---VMTQIIRGQPVSTAVSAPNT 2098
 Db 1784 LKNTPSEGOQNTGL-KNTPSEGOQNTGLKNTPNEROQNTGLKNAANKGOQNTGLKNTPNE 1842
 QY 2099 -----VSTP--GOKSLTSATSTSNIOSSASQPPRPOQGYKLTMAQLTQLTGHGNO 2150
 Db 1843 GOQNTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNAANKQ---QNT 1899
 QY 2151 GLTVVIOGQQTGLOLIPQCVTVLPBGQOLMAAMPNGTVQRFPLPATITAST 2210
 Db 1900 GLKNTPSGGQONT-----GLKNTPNEGQ---NTGLKNTPSEGOQ 1936
 QY 2211 TTTVSTTAAGTGEOROSKLSPOMVHOD---KTLPPAOSSSVG---PAKAOPOTA-- 2260
 Db 1937 NTGLKNTPSEG---QONTGLKNTPNEROQNTGLKNTPSEGOQNTGLKNTPSEGQNTGLK 1993
 QY 2261 -QPSARPOPTOPQSPAPPEVOT---QPEVOTQTTVSSHPSAOPTHAOSSKPOVAQ 2315
 Db 1994 NTPNEGQONTGLKNTPSEGOQNTGLKNAANKGOQNTGLKNTPNEGQONTGLKNTPSEGOQ 2053
 QY 2316 SQPOSNV--QOSPVRYQ--SPQOTRIRPSTBSQLSPGOQSOVQTTISOPIPITPHITSLQ- 2371
 Db 2054 NTGLKNTPSEGOQNTGLKNTPSEGOQNTGLKNAANKGOQNTGLKNT--PNEROQNTGLKN 2111
 QY 2372 IPSOGQPSOPQVOSSQTLSGGOTLQNVSVSPSRPOLQIQOPQVIAVPQLQOQVY 2431
 Db 2112 TPNEGQ-----QNTGLKNTPSEGQO--NTGLKNTPSE-----GOQNTGLKNTPSEGOQNTG 2160
 QY 2432 LSOIQSQVVAQIOAQSGVPQIKLQLPQIOQSSAVQ---THQIQNVVTVQNASVOEQL 2488
 Db 2161 LKNTPSE-----QOPNTGLKNTPNEGQONTGLKNTPSEGOQNTGLKNAAN----- 2205
 QY 2489 QRVQOLRQOQKQKQOQIEINVNTPSKLLIKVEYITQKOVYMKHNAVIEHLKOKSKMTPAE 2548
 Db 2206 -----KGOQNTGLK-NTPSE-----GOQNTGLKN-----TPSE 2232
 QY 2549 REENORMIWCNOVMYIIDLKIDKEKQAKKRRRESEVQKRSK--QNAFKLSALLFKHKE 2607
 Db 2233 GOQNTGL-----KNANKGOQNTGLKNTPSEGOQNTGL----- 2265
 QY 2608 QLRARILKKRALDLQIEVOEELKRDJLIKKEKDLMLAQATAVAAPCPVTPVLPA 2667
 Db 2266 -----KNTPSEGOQ--NTGLKNAANKGOQNT-----TGKNTP 2295
 QY 2668 PAPPPSP-----PPPGVOHTGLSTPLTPVASQKRRREKDSKSKKKMISTTSKET 2723
 Db 2296 SEGQPNITGLKNTPNEGQONTGLKNTP-----SEGQONTGLKNTPNEGQONTGLKNTPSEG 2350
 QY 2724 KKDFTL 2729
 Db 2351 QONTGL 2356

Search completed: November 20, 2002, 16:31:35
 Job time : 142.041 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:33:37 ; Search time 13.288 Seconds

(without alignments)
3426.253 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAEETDSE.....KLKGFKASRSHNNKQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1385	9.3	328	US-09-925-297-816	Sequence 816, Appl
2	982.5	6.6	238	US-08-729-835-80	Sequence 80, Appl
3	443	3.0	2344	US-09-815-242-12713	Sequence 12713, A
4	438	2.9	3256	US-09-919-172-98	Sequence 98, Appl
5	432	2.9	2665	US-09-864-761-34248	Sequence 34248, A
6	402	2.7	4019	US-09-738-973-425	Sequence 425, Appl
7	401.5	2.7	1367	US-09-801-368-108	Sequence 108, Appl
8	391	2.6	5179	US-09-922-217-1068	Sequence 1068, Ap
9	391	2.6	5179	US-09-833-263-1068	Sequence 1068, Ap
10	389.5	2.6	2478	US-09-815-242-5816	Sequence 5816, A
11	389.5	2.6	2478	US-09-815-242-12967	Sequence 12967, A
12	361.5	2.4	2843	US-08-681-219-32	Sequence 32, Appl
13	354.5	2.4	1400	US-09-764-176-7	Sequence 7, Appl
14	342	2.3	2368	US-09-815-242-5635	Sequence 5635, Appl
15	342	2.3	2368	US-09-815-242-12389	Sequence 12389, A
16	339	2.3	6281	US-09-815-242-12996	Sequence 12996, A
17	329.5	2.2	5795	US-09-815-242-12610	Sequence 12610, A
18	324	2.2	1781	US-09-738-877-3	Sequence 3, Appl
19	318	2.1	1596	US-09-902-432-4	Sequence 4, Appl

20	317	2.1	1363	US-10-124-557-52	Sequence 52, Appl
21	315	2.1	2063	US-09-735-367B-2	Sequence 2, Appl
22	312.5	2.1	1140	US-10-124-557-104	Sequence 104, Appl
23	312.5	2.1	1404	US-10-124-557-2	Sequence 2, Appl
24	312.5	2.1	1404	US-10-124-557-62	Sequence 62, Appl
25	312	2.1	2005	US-09-735-367B-3	Sequence 3, Appl
26	311.5	2.1	1049	US-10-124-557-18	Sequence 18, Appl
27	311.5	2.1	1313	US-10-124-557-142	Sequence 142, Appl
28	311.5	2.1	1354	US-10-124-557-48	Sequence 48, Appl
29	309	2.1	2441	US-10-109-886-8	Sequence 8, Appl
30	306	2.0	957	US-09-922-217-1065	Sequence 1065, Ap
31	306	2.0	957	US-09-833-263-1065	Sequence 1065, Ap
32	305	2.0	1320	US-10-124-557-46	Sequence 46, Appl
33	305	2.0	1320	US-10-124-557-60	Sequence 60, Appl
34	305	2.0	1361	US-10-124-557-40	Sequence 40, Appl
35	304	2.0	1314	US-10-124-557-50	Sequence 50, Appl
36	301.5	2.0	3158	US-09-815-242-12611	Sequence 12611, A
37	301	2.0	1038	US-10-124-557-74	Sequence 74, Appl
38	301	2.0	1370	US-10-124-557-44	Sequence 44, Appl
39	301	2.0	1311	US-10-124-557-42	Sequence 42, Appl
40	300	2.0	1022	US-10-124-557-84	Sequence 84, Appl
41	296	2.0	1537	US-09-801-368-104	Sequence 104, Appl
42	293.5	2.0	941	US-10-124-557-14	Sequence 14, Appl
43	293	2.0	2442	US-10-109-886-10	Sequence 10, Appl
44	292	2.0	2785	US-09-801-574-8	Sequence 8, Appl
45	290	1.9	2086	US-09-815-242-5639	Sequence 5639, Ap

ALIGNMENTS

RESULT 1
US-09-925-297-816
Sequence 816, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA105
CURRENT APPLICATION NUMBER: US/09/925,297
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05989
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (178)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (183)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (269)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (286)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-297-816

Query Match	9.3%;	Score 1385;	DB 10;	Length 328;
Best Local Similarity	66.4%;	Pred. No. 2.9e-53;		
Matches 286; Conservative	0;	Mismatches 17;	Indels 128;	Gaps 2

QY	406	NKSLTAAANEIELESIAKKGGDIDNVKSP--EFTKXKNTENDSKDAENRFEPOSJL	463
Db	12	NVSLSGSTNEIELESITAKKGDDIDNVKSPSTGEETEKDKNETENDSKDAENRFEPOSJL	71
QY	464	EKSDSDKTPDDDEPQGSSEEPTEVGDKGNSVSANLGDNTTNAATSEETSPSEGRPYGCLS	523
Db	72	EKSDSDKTPDDDEPQGSSE-----	90
QY	524	ETPDSSMAKKAASELPQDVPEEPNKTCSSTNTSATTTISIOENLNSNSSELSNOSSE	563
Db	91	-----	90
QY	584	SAKAADDPENGERSHNPVasioEIVGPTSEKSTGEELSEPGAGKAGSGSTRITRLRN	643
Db	91	-----VDFEKSXNGHLSFSPAGKAGSGSTIITRLRN	125
QY	644	PDSKLSLOKSOQVAAAHAANEANKLFKEGKEVLVYNSOGIEISRLSTKKEVIMKGINNYFKL	703
Db	126	PDSKLSLOKSOQVAAAHAANEANKLFKEGKEVLVYNSOGIEISRLSTKXKXVMKGXINNYKKL	185
QY	704	GOEKYVYVYNOYSTNSFALNKHONRHDHKKRRHLNHPCLTPAGGEPKMGSHVSGSKVLT	763
Db	186	GOEKYVYVYNOYSTNSFALNKHONRHDHKKRRHLNHPCLTPAGGEPKMGSHVSGSKVLT	245
QY	764	ISTRLRTITOLENNIPSSFFHPNASHANWIKAYOWMSKPRFALALILECAYAPVYM	823
Db	246	ISTRLRTITOLENNIPSSFFHPNASHANWIKAYOWMSKXPRFALALILECAYAPVYM	305
QY	824	LPIWREFLGHT 834	
Db	306	LPIWRESLGHT 316	

RESULT 2
1960-77

```

US-09-129-835-80
? Sequence 80, Application US/09729835
? Patent No. US20010016647A1
? GENERAL INFORMATION:
? APPLICANT: Ruben et al.
? TITLE OF INVENTION: 29 Human Secreted Proteins
? FILE REFERENCE: P2015P1
? CURRENT APPLICATION NUMBER: US/09/729,835
? CURRENT FILING DATE: 2000-12-06
? PRIOR APPLICATION NUMBER: 09/257,179
? PRIOR FILING DATE: 1999-02-25
? PRIOR APPLICATION NUMBER: 60/056,270
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,271
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,247
? PRIOR FILING DATE: 1997-08-29
? PRIOR APPLICATION NUMBER: 60/056,073
? PRIOR FILING DATE: 1997-08-29
? NUMBER OF SEQ ID NOS: 128
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 80
? LENGTH: 238
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-729-835-80

```

Query Match	6.6%;	Score 982.5;	DB 10;	Length 238;
Best Local Similarity	78.8%;	Pred. No. 4.2e-36;		
Matches 189;	Conservative	7;	Mismatches 19;	Indels 25;
				Gaps 3;

QY 2689 P1LEVASQKRRKEEK-----DSSSKSKKKKMISTTSKETKK-----DT 21227
|| | | | | : : : : | | | |

Db 3 PTPR-----KTPYDESKFYIGCDLCTNMYHGECVGITEKEAKKMDVYICNDCKRAQEGSSE 58

Qy	2728	KLICKCTPDESEKEYTGCRCOMNHWGRCVGLIOSEALIDEYVCPQOCSSTEDMAYLT	2787
		::::::::::::::::	
Db	59	ELVICICTPDESOFTYIGCDRCOMNHWGRCVGLIOSEALIDEYVCPQOCSSTEDMAYLT	118
Qy	2788	PLETEKDEGLRVLRSLQAHKMMAPLEPVDPNDAADYGVATKEPMDLATMEERQYRY	2847
Db	119	PLETEKDEGLRVLRSLQAHKMMAPLEPVDPNDAADYGVATKEPMDLATMEERQYRY	178
Qy	2848	EKLETFEADMTKTLFDCRCRYNPSDSPFYOCAEVLESFFYQKLKGFKASRSHNNKLOSTAS	2907
Db	179	EKLETFEADMTKTLFDCRCRYNPSDSPFYOCAEVLESFFYQKLKGFKASRSHNNKLOSTAS	238

RESULT 3

```

US-09-815-242-12713
? Sequence 12713, Application US/09815242
? Patent No. US20020051659A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Kai L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of ESRs
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0.
? SEQ ID NO 12713
? LENGTH: 2344
? TYPE: PRF
? ORGANISM: Staphylococcus aureus
US-09-815-242-12713

```

Query Match	3.0%;	Score 443;	DB 10;	Length 2344;
Best Local Similarity	18.7%;	Pred. No. 8.4e-12;		
Matches 437;	Conservative 308;	Mismatches 990;	Indels 598;	Gaps 76;

[illegible]

Db 238 SRLAMSTFASAA-----TTTALTANTTVAKNDIKOYMTTSGAATDOSTGVYT 286
Qy 621 LSESPAGAGA-SGSTRITRLRNPD SKLSQOAAAHAEANKLFEGEK---EVLVY 676
Db 287 LPODPTSGKALITGLTRI-----DSN---KSFHSGKVLNGLNKEGNGNGDGIIFA 335
Qy 677 NSQGEISRLSTKEVIMKNNINNYFKLGOEGKYRYHNOYSTNSFLNK-----725
Db 336 FSPGVGLGTNGAAVIGLSNAFGF---KIDTYHNTSTPNSSAKAKADPSNVAGGA 391
Qy 726-----HQHREDHDKRHLAKFCLPAG-----EKKMGVSGSVLVT 764
Db 392 FGAFTYTDYGAAYTTSSTADN---AAKLVOPTNNTFODEDTINYGDD---TKMTY 444
Qy 765 STLRILITOLENNIPSSFPHPNASHRAMIKAVOMCSKPREFALALA-----812
Db 445-----KYAGQITWIRNISDMIAK---SGTJNLSLMTASTGATNLQOY 484
Qy 813-----ILECAVKPVYMLPIW-----REBLGH-----TRLHRTSTIE 843
Db 485 QGTEFEYTESAVTOVRYVYVTTGKDIIPKITYSGNDQVYVITDNOQSALTAKGNYTSDY 544
Qy 844 REKEKVKKEKKEKOEETMOQATWYKYTFEPVKH---OYWKQGEYR-----888
Db 545 SGTASTYNDTKTKVKTNAQOSYTT---YFTDYKAPTYVGNQTIIEYKTMNPVLVTTDN 602
Qy 889-----VTGY-GGWSMISTKTHVRFVFLPGNTNV-----NYRKSLEGTKNMDE 931
Db 603 GGTGYVNTVGLPGLSGYSDATNSIIGPTPKIGQSTVYVSTDQANNKSTTFTTINVDI 662
Qy 932 NM-----DESDKRKCSPPKKIKIEPDSEK---DEVKGSDAKADONEMDISKITEK 981
Db 663 TAPVYTPIDKSEVFPSPISPIINATODNSGNNAVNTVGLPSGLTFEDSTNNITISGPTN 722
Qy 982 KQODVKEILDSS--DKPCKEPEMEYDDMKTESHVNQCESOVADVNVSEGHLTYSK 1039
Db 723 IGTSTITTYTSDASGKNTTTRTKEYEYTRNSMSDVSSTSGSTQOSQSVSTSKA-----D 775
Qy 1040 KTKKSKLDGLERIKOFTLEKORLEKIGKIGIKGTSTNSKNLSESPVITTKAK 1099
Db 776 SOSASTSTSG---SMTSTASTSTKSTSVLSDSVASKSLSTSESNVSSTSTSLVN 831
Qy 1100 ECGQSMNOEOSPNA-----NNDOPEDLIQCCSOSDSVLRMS-----DPSHTNKLX 1148
Db 832 SOSVSSMSGVSKSTSLSDFLSNSSTSEKSESVSTSDSLTSTSLSDSVMSSTSGSL 891
Qy 1149 PKDRVLDVVISPEPKCKONSIENDI--EKKVSDLASRGOEPTKTKGNDFITDCK 1206
Db 892 SKQSLSLSTSDASSTOSVSDSTNSISTSELSSEGSTSEISISNSISNSVASTSK 951
Qy 1207 LASADDIGTLICKNKPLIOEE-----SPTIVSSSKSALHSSVPKSTNDRDAP--L 1256
Db 952 LBSQSTISLSTSDSKSMSTSELSLSTSDSVSGSLVAGSQSTSLSTSDSMSTSEMI 1011
Qy 1257 SRAMDEGLIGCD-----SESNSTLENSSDTVSIODSSEDMIVONSSESIS 1303
Db 1012 SOSMSTSGSLASDSKSMSSVSSMSTSGSTSESLSDSISTSDSKSLSTISQSGST 1071
Qy 1304 EDPFRREDQVEVLEPLKCLVGESENGCEDRLPVKGTPLANKKPKSQOKLEERPYNKS 1363
Db 1072 STSTSTSSSVRME--SGSTSGSMSTSQSDSTIS--TSFS 1108
Qy 1364 DOIKLNTTKKNNENRESEKKGORTSTFOINGKONPKIYLKGECLKETESRVSAGNV 1423
Db 1109 D-----STSDSKASASTASESISTOSYST-----STSGSISTST 1141
Qy 1424 EKVNNINKIIE-NDIKSLTVKES-----AIRPFLNG---DVIMEDFERNRS 1467
Db 1142 SLTSTNSERTSTMSDSTSLSTSESDSTSDSTSDSISEALISGSTSLSTSESNSTSD 1201
Qy 1468 SETK-----SHLSSDAE-----GNVROSLLETPTKSADSQTTPPSACPESNS 1514
Db 1202 SESKSASAFLESLSESTSESTSESLSGSTSDSTSLSDSNSESGSTSTSL--SNSTSGSAS 1260

Qy 1515 VNQVEDMEIETSEVKKYVTSPTISEESNLSDNFIDENGLPIKNENNGE-----1565
Db 1261 ISTSTSGASASTVK---SESVSTSLSTSTSLSDSTSLSTSLSTSDSGSKNSLSASM 1317
Qy 1566-----SKRVTYTEVTMTSTVATESK---IYIKEREKDK-----QIVVSS 1603
Db 1318 STSDSISTRSSESLASTSLSGSTSESESGSTSSSEKSDSTMSLSMGSOSTSGSTSVST 1377
Qy 1604 TENCASVTY-----TTTTVTYKLSPTSGSVDIISVKEQS 1640
Db 1378 SELSLSDSTSLSLASMANOSGVDSNASASOASASTSTSTSESDOSTSTSYTSQSTQSE 1437
Qy 1641 KTVVTTVTDSL-----TTTGTLVTSMTVSKYSTRDVKYKLMKFSRPPKTRSGTALPSY 1695
Db 1438 STSTSLSDSTSLSKSTSGOSTSTASLSGSESDSQSISTSTSEKSESTSTSLSD 1497
Qy 1696 RKFTVSTKKSITVLEPNDLKLARKGIGREVEYFNYNNAKPAIDIMPTSPPTFGITMR 1755
Db 1498 STSTNSGASASTYL-----LSNBSASASESDSSSTSLSDSTASAMQSESDS-----1544
Qy 1756 YRLQTVKSLAGVSLMLRLMASLRMDMAKVPFGGSGTPTETSETEITTEILKRRDVG 1815
Db 1545 --OSTSTSLNSOOSTSTIRMTIASESVSESTSESGSTSESTSESDSTSLSDSQST 1602
Qy 1816 PVGIREPEYCIKILICIGPET-----PREPTPORKGLRSALRPKRPPEKQTPYI 1869
Db 1603 -----RSTASGASASTSTSDSRSTASSTSMRTSLDQSMLSLSTSTSV 1651
Qy 1870 IETWAELELELMEITAFERKEKKAQAVEQAKRLEQAKTYVATSTSTSTSTST 1929
Db 1652 SDS-----TSLSDSVSDSTSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1689
Qy 1930 ISPAQKVNAPISG-----SVTTGKVMVLTKVGS 1959
Db 1690 STASEVMAASISDSOSMSRVNDSVDSSESVSNSESPKSMGSTSVSDSGSLSTSLRK 1749
Qy 1960 PATVTFQOKNFFHQTEPATVWKQOSNGSVYVQOKVGLIIPSTSTSTSTSTSTSTST 2019
Db 1750 SESVSEISLISGQSMDSVSTSDSSLSLAVSTQORSESESVSESDSLSDSKSTGSTST 1809
Qy 2020 V-TIRPTSGSG-----TMSQVITGPQIRPMGVITTPLOQST 2059
Db 1810 SGSLSTSTSLSGSESVSESSLSLSDSISMSDSTSTSDSLSGSISLSDSTSLSDSLD 1869
Qy 2060 LGAIRTRPVMVQPGAPQOVMTOIIRGQPVSTAVSAPNTVS-----STPGQKSLTS 2110
Db 1870 -SKSLSSQSM---SGSESTSVSDSQSSSTSTNSQFDSMISASDSKMSSTSDSSISG 1925
Qy 2111 ATSTNSIQSASQPPRPOQOQVYKLTMAOLITOLQHGNGGLTVVIQOGQOTTGOLQIIP 2170
Db 1926 SNSTSTSLSTSDS---MSGSVSVTS--TSLSDSLSGS--ISVSDSSSTSESLIS 1974
Qy 2171 QGVTVLPBGQOQLMQAMPNGVQRFLEPPLATATTAATTTTSTVSTAAGTEBOROKU 2230
Db 1975 -----DSMAQOSQSTSTASAGSLSTISLSMSASAGTLISOSTSVSTST-----L 2017
Qy 2231 SPQMOVHOKTILPPAASSSVGPAKAPQTAQPSARQPOQTQPOSAPQEVQTPVQOTQ 2290
Db 2018 STSDSISDSTISISISGQSAVESESTSDSTISDSSELSLSTSGSTSTSTSESLSTSN 2077
Qy 2291 TVSSHVPEAOPTHAQSFPQVAQAQOSMPOVNOGQSPVRYQSPQRIRPSTPSQSLPQ 2350
Db 2078 SGSTSV-SESLSTSGSGS-----TSYSDS-----STSSSLSTSG 2111
Qy 2351 OSQOVOTTSQPIPIQHTSLQIPSGQPOQPOQVOSQTOQLSSGQTLNOVSVS 2403
Db 2112 STSVSDSTSMSESNASISMSQISGSTSGSTSTSTSESLSMGSGTHNSTSVS 2164


```

; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

```

```

Query Match      2.9%  Score 438; DB 10; Length 3256;
Best Local Similarity 18.3%; Pred. No. 2e-11;
Matches 560; Conservative 418; Mismatches 1068; Indels 1010; Gaps 143;

```

```

QY 300 VPGVTDCAVEIQNKPKYIRHEPIGYDSRRKRYWFLNRLIIEEDTENENEKKIWM----- 354
DB 221 VP--TTCOLDNSKNE-----SPFKIYESVAKKELDVSQKENVLYQYCKSG 265
QY 355 -----YSKQVLAELIDCLDKDYEAELCKLLEEMREIHHMDITEDLTAKAGSNKSEFL 410
DB 266 LQTDYAIATEKESADGLQ-----GETOL-----VSRRKSRRKSGSGSHAVA 304
QY 411 AAAN-EEIEESTIRAKKGDIDINVKSPET-----EKDKNET-----ENDSKADEKNR 455
DB 305 EFAESPEDLDONKKGKRDVSVOTPSKAVAGASPLYPRAKMKTPVOYSQOQNSPQKHKK 364
QY 456 EEF-----EDOSLEKD---SDKTPDDDEQKSEEPTEVGDKNSSVS---ANLGDNTN 504
DB 365 DLYTTGREGSVNLGKSGEFGAGDKTLTPRKLTNRNTPAKVEDADAATKPENLSSKTRG 424
QY 505 A--TSEBTSE---GRSPVGCLET-----PDSSNAEK-----KYASELPQDYPE 546
DB 425 SIPTDVEVLPTEIEIHNEPLTLMLQVERKIQKDSLKPEKLGTTAGQMSGLPLSSV 484
QY 547 EPNKTESSNTS-----ATTTSIQPNL---ENSSSSELNSSOSESAKA----- 588
DB 485 DINPFDSINSEGCIPLAKRRVSPFGHLRPELDENLPRPTPLKGEAAPTKRKSLVWHTP 544
QY 589 -----DDPE-NGERES---HTPVSIQEEIVG-----DFTSEKSTGELESPPG 626
DB 545 PVLKTIIEKQPPQSGKQESSEIHVEVKAQSLVISPAPSPRKTVPVADRRRRSCKTAPA 604
QY 627 AGKGA-----GSTRIITFLRNPDS-----KLSQLKSOQVAAAHAAEKLKEGEEVL 674
DB 605 SSSKSQTEVPKRGGERVATLQKRVSLISQHDILQICSKRRSGASEAN-----LI 656
QY 675 VVNSQGEISRLSTRK---EYIMKG---NINNYFKLQEGKYRV--VHNQYSTN-----S 720
DB 657 VASMSWDVYVLGAKQOTQTKVIKKGQPSMMKRRRPRATPKKPYGVEHVSQSTIGHANSPCT 716
QY 721 FALNKHOHREDHDKRR--HLAHKFCCLTPAGEFK-----WNGSVHSGKVLITLTRL 769
DB 717 IILGKAHTEKVEHVPARPYRLNMFISNQKDFEKEDLSGLIEMFKTPVKEQPQLT-STCHI 775
QY 770 TITQLNNIIPSSP-----FHPNMASHRAWIKAVQCSK--PHEFALALALILEAVK 819
DB 776 AINSSENLGKQFGTDSGEEPLLPTSESGVGFVFSQAQAAKQPSD-----KCSAS 827
QY 820 PVVMLPIMREFLGHNT-----RLHMTSIIERE-----EKEK 849
DB 828 P-----PLRQCIENGNVAKTPRNTYKMTSLFETKSDTETEPSTYVTVNRSGRSTEFNR 883

```

```

QY 850 VKR--KEKKQEEETM-----QOATWVKKYTFPVKQHVWKQKGEYRVGTGYSWSMT 898
DB 884 IQKLVPYSKSEETNTELVCEILKRGQATLLQ-----RREGMKEL----- 925
QY 899 SKTHVYRFVFKLPQNTNVNRYKSLGKTKNMDEMDESDK-----RKCS--RSPKTKIE 951
DB 926 -----ERPFEI---YKENIETKEN--DEKMKAMKRSRTWGOCACAPMSDLTDLKSL 970
QY 952 PDSE--KDEYKG-----SQAAGADQENEMDISKITKKNQDVKELLDSQDPCKE--EP 1002
DB 971 PDLKMKDTARQONLQTDQHAAPKSEKTKTKM-----PCOSLQ 1012
QY 1003 MEYDDMKTESHVNCOESSQOVYVNVSE-----GFHLRTS-----Y 1038
DB 1013 -----EPINTPTHTKQQLKASLGKVGVEELLAVGKFTRTSGETHHREPDAGGKSIRTF 1068
QY 1039 KKKTK-----SSKLIDGLERRIKQFTLEEKORLEKIKLEGIGIGTSTNSSKNISESP 1093
DB 1069 KESPKQILDEAAAVTGM--KKWPRTPKEEAQSLEDL---AGFKELQTPQPSSESMTEK 1123
QY 1094 VITKAKGCSQDSMRQPSQSNANNDOPE-DLIGCSQSDSVLRMSDPSTTKLYPKDR 1152
DB 1124 T---TKIACSPPEESVDYTTSTQWPKRSLKRAVDEEFLALRKLTSPSAGKAMLTPKPA 1180
QY 1153 VLDDVSI-----RSPETKCPKQNSIEN-----DIEEKVSDL 1183
DB 1181 GDEKDIKAFMGTPVQKLDLAGTLPGSKROLQTPKEKAQLELDAGKELFQTPGHTEEL 1240
QY 1184 ASRGQ-----EPTKSKTKGNDPFIQDSKLASADDIGTLI--CKMKKPLIOE 1227
DB 1241 VAAKGTTKIPCDSPQSDPVDPTPTSTQKQ-----PKRSIKRAVEGGLACRNIMP--- 1290
QY 1228 ESDTIYSSKSLSHSVKSTNDNR-----ATPLSRAMDEGKLGCDSESNSTLSSSDT 1282
DB 1291 -----SACKAMHTPKRPSGEEKDIIIFVGTVOK-----DLTENLIGS 1329
QY 1283 VSIQDSSEDMIVQNSNESISEQFRTREDQVEVLEPLKCELYSGESTG--NCDRLPYKGT 1341
DB 1330 KRRPQTPKEEAQALEDTLGTKELFQTPGHTEEA-----VAAKGTTKMKPCESSPESAD 1382
QY 1342 EANGKRQSQOKLEERPVNKCSDQIKLNTDKNNENRESEKKGQORTSPQINGDNRP 1401
DB 1383 TPTSTRQOPTLEKRVQVQELSALK-----KLTQVSGETHYDKVGGEDKS 1430
QY 1402 KIYLKGECLKEISESRVANSV--EPKVNINIKIIPENDISLTVKESAIKPFINGVIME 1460
DB 1431 INAFREAKQKLDPAASVYTSKSRHPKTK--EKQAPLEDLAGW--KELFQPR-----VCT 1480
QY 1461 DFNERNSETKSHLSSDAEGVYRQSLPTPSTKESDSTQTTTPSASCPESNVQVED 1520
DB 1481 DKPTTHEKTKIACRSQPD-----PVDTPSSKQSKRSRLKVD 1519
QY 1521 MEIETSEVKKVTSS-----PITSEESNLSPNDFIDENGLPIK--NEWNVGSEKR 1568
DB 1520 VEEEFALRKRTPSAGKAMHTPKPAVSGEKNIYA--FM--GTPVQKLDLTELMTGSKR 1574
QY 1569 KTVITEVNTMTSTVA-----TESKTIYKVEGQKQTVVSTENCASSTVYTTT 1616
DB 1575 LQTPKEAQAQLELDAGKELFQTRGHTESMT---NDKTAKVACKSSQPDLDKNPASSKR 1631
QY 1617 TTVTKLSTPSTGSDVLIISKEQSKTYVTTVDLSLTGTLVTSMTYKESVSTNDKYK 1676
DB 1632 RLKTSLG--KVGVKEELIAGVLTQSGETHHTHEPTGCGSKMAFMESPKOILDSAS 1689
QY 1677 LMKFSRPKTKRSGTA-----LPSYRK--FVTKSTKKSIFVLPNDL----- 1715
DB 1690 LTGSKRQRLTPPKKSEVPEDLAGFIELFQTPSTHKSMSINEKTKKSYRASQPDVLDTPT 1749
QY 1716 -----KTLARKGIREVPYFNANAPALDIWYPSBPFPGITWRYRLQTVKSLAGVSL 1769
DB 1750 SSKPQKRSRLKADTE-----EFLAFRQOTPSAG-----KAMHTPKPAVAGEEK 1793
QY 1770 MLR-LIMASLRWDDMAKAVPGGGSTRTETSETETITTEIIRRDVGPYGIR----- 1820

```


Db 1794 DINTFGLTGVOKIDOPGNLP--GSMNRRLQTRKKAQALELTL-----GFRELFQTPCT 1844
QY 1821 -----FEYCIKRLIC--PLGVP--ETPKEMPTPOKGLRRS-----ALRPKRP----- 1859
Db 1845 DNETDEKTTKTLKSCPSQSDPADPTNTKORPKSLKADVEEFLAERKLTLPASGAKM 1904
QY 1860 EMPKQGTGVPIIETWVAEELEIMEIRAF-----AERVE-----KEKAQAVE 1900
Db 1905 HHPKAA-----VGEER-----DINTFVGTVPYKXLDLGNLPKSRKRPOTPKKKKALE 1952
QY 1901 QOAKKRLBQOKPVIATSTTSPTSTTSPAKQVNAVAPISGVTGTMAVLTAVGSP 1960
Db 1953 DLAGEFELQTPGHTESMTDKITEVSCSPQDPVKTTP-----TSSKORLIKISIGK- 2005
QY 1961 AATVFOGNNHFOTFATWVKQGSNSGVVQOQKVL--GIIPSTGTSCQTFSPQRTA 2018
Db 2006 -----GVKKEVLPVGLTQTSKTYOT----- 2028
QY 2019 TVTIRPNTSGSGGT-----TSNSQVI-----TGPOIRPGMTIIRTP-----LQOSTLG 2061
Db 2029 -----HNETAGDGKSIKAFKESAKQMLDPANYGTMERWP-----KTPKEAOSLEDLAGF 2079
QY 2062 KAIIRTPVAVQEPAPQOQVMTQIIRGQPVSTAVAPNTVSTPGQKSLTSTSTNIOSSA 2121
Db 2080 KELFOQTPDHTHEESTTDDKTKTKI-----ACKSPPESSMDPTST----- 2117
QY 2122 SOPPRQOGQVOKLT--MAOITOLTOGHGNGQLTVVIQGGGQVITGLOLILP-----OGVTV 2175
Db 2118 RRRPKTPLGKRDIVEELSLAKQILTO-----THTDKVGGDEGGINV 2159
QY 2176 LBPFGQOLM-----QAAMPNGVVO-----RFLF--TPLATVATATSTTTTV 2215
Db 2160 FRETAOKLDPAASVYSGKQPRTPKGAQRLPDLAGLKLFTQPLCTODKPTTHEKTTKI 2219
QY 2216 SITTA-----AGTGE--QROSKULS--POMOVHODKTLPPAOSSSVAPAKAPQTA----- 2260
Db 2220 ACSRPQDPVGPPTTIFKPKQSKSLKADVEEESLALRKRTPSVKAMDPTPKPAGSGEDKM 2279
QY 2261 -----QPSARPOQOTQPOSAPQEPVOTQPEVOTQTVSS-----HVPSEAKPT 2303
Db 2280 KAFMGTPVOKLDPGLPGLPSGRMPQTPKE-----KAQLEDLAGKELPOTPGDKPT 2332
QY 2304 HAQSKPOVAASQPOSNOVQOSPVAVOSPQTRIRPS-----TPS- 2344
Db 2333 -TDEKTTKLACKS-PQ-----PDPVDPNPASTKQRPKRLKADVEEFLALRKRTPSA 2383
QY 2345 -----QLSPGQOSQVOTTSQPIPIOPHTSL-QIP-SQOGPOSOP----- 2382
Db 2384 GRAMDTPKPAVSPDEKNINTEFETPVOKLIDLGNLPKSKRPQTPKKEALEDLVGFKEL 2443
QY 2383 -QVQSTQTLSSQOTLNOVSVSSPSRPQIQOPQOVIAVP-----QLQOQOVLSQI-- 2435
Db 2444 FQTPGHTESMTDKITEVSCSPQESFTSSKQRLTPIPLVKVMDKEPLAVSLTR 2503
QY 2436 QSOQVNAQIAOOSQVPOQIKL--QLPITQIOQSSAVOTHOIQNVVTVQAASVQOLORVQO 2493
Db 2504 TSETQOTHTHEPTGDSKSIKAFKESPKQILDPAA-----SVTSRRLQTRKKEKARA 2555
QY 2494 LBPQOQKK-----QOOIETINV-----TPSKLLIKVEIITQOVVM 2529
Db 2556 LLEDLVDFKELFASAPGHTESMTDKNTKIPCKSPPELTDATSTKRCPTKRL--KREVK 2613
QY 2530 KHAQVIEHLKOKKASMT-----PAERENQRMIVCNOVMYIILDKIDKEQAKKRRRE 2583
Db 2614 EELSAVERLQTSVGSSTHTHKEPASGDEGIKVL-----KORAKKPNP 2656
QY 2584 ESYEOKRSKONA-----TKLSALLFKHKEQLRA-----ELKK 2616
Db 2657 VEPEPSRRRPRAPKAKQPLEDLAGFTLESETSGHQSGLTACKATKIPCESPPLVVDFT 2716
QY 2617 RALLDQLOIEVQEEELKRDIKKEKDQMLQAO-----AT 2651

Db 2717 TASTKHLRLTRVOK-----VQVKEPSAVKFTQTSGETTDADKEPAGEDKIGALKESAK 2771
QY 2652 AVAAPCPVTPVLPAPAPPPSP-----PPGVQHTGLSTPTPLPVASQKRRREE 2703
Db 2772 QTPAPASVYSGRRRPRAPRESAQALIEDLAGFKDPAAGHTESMTDKTKTKICKSPEL 2831
QY 2704 KDSSSKSKK-----KMISTSKETKDTLYCICKTPYDESK 2741
Db 2832 EDATSSKRRPRPRAQKVEYKEELLAVGKLTQTSGETTHD-----KEPVGEK 2880

RESULT 5
US-09-864-761-34248
; Sequence 34248, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14


```

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; OTHER INFORMATION: EST_HUMAN HIT: A0117052.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALU0 3.00e-10
US-09-864-761-34248

```

```

Query Match      2.9%; Score 432; DB 10; Length 2665;
Best Local Similarity 19.1%; Pred. No. 2.9e-11;
Matches 547; Conservative 397; Mismatches 1112; Indels 810; Gaps 131;

```

```

QY 7 EEDGDAEETQDSEDD-----EDMEEDDDDDYPEEMEDDDDDASYCTSSFRSHS 59
DB 170 KNEITDKERTFDEPERERRLIRKEKVEKDKTKDQ----- 205
QY 60 TYSTPRRRPRVHRPSPLF-----EKDIPLEFPKS-----SEDLAVPNEHT 104
DB 206 -----KRGKGVHSPSSQSEETDQENERSQSP-EKPRSCNKLREKADKEGIAKNREL 257
QY 105 MNVAIAYEVNFTVRLSPRFEEDFCALVSOQCTTMAEMHVLLKAVLEEDTSNT 164
DB 258 MCVVILTRVKEKGVIDHTPV-----EKAKALMDITYKSSALDCKLOYST 305
QY 165 TEGPADIK--DSVNSTLYFDGM-TMEVILRVYCESD--KEYHNVLPYQEAADYPPY 218
DB 306 EPARKSDLSKLESYRMKPKKEGLSHSEYV-----EKEGRLKARKHLKPEQPAD-----GVS 357
QY 219 ENKIKVLOFLVDQFLTNIAREELMSF---GVIOYDHCHVCNKLDDLCCEFCSAVYHL 275
DB 358 AVLEKLEAKRRFPADSNLAEKQKPEVKKSSPEMED-AVLSKOPDVSSREVILLREG 416
QY 276 ECVKPPPL--EEVDEMOCEVCVAHVPGVTDCAVIOKKNPYIRHEPIGYDSRRKYM 333
DB 417 EARRKPVKREILKRESKIKLIDLNTVYASPKDC-QELAS-----ISVGSGRSSD 466
QY 334 LNRRL--IIEDEPENEKIKIYSTRKVLAEILDLCKDYWEALCKIIEEM--REET 388
DB 467 LQARLELAGESEVNOEVOSKRPISKPOLKOL-----QVLDGDPGPREYV 512
QY 389 HR-HMDITEDLTNRKAGSNKSFLLAANEETLESIRAKG-DIDNVKS--PEETEDKNE 443
DB 513 RKVYCLRDTPPRKSGQEKSHSVNTEKTI-----GIDIDHTOSYRKOMEQSRKQO 564
QY 444 TENDSDAE-----KNREEFEDQSL-----EKSDDKTPDDDEQKSEPTFVG----- 488
DB 565 MEMEIKASEKFGSPKRDVDEYERSLVHEVGRKPPQDVTDSPSKKRRMDHVDICTKR 624
QY 489 DKGNVSANLGDNTNATSEET--SPS-----EGRSYGC--LSETPDSSMAEKVYA 537
DB 625 ERNRRSSROISD-----SERTGSPSVRHSFHEDEDPISPRLLSVKSPVYDK-- 676
QY 538 SELPQD--VPEEPNK--TCESSTATTTSTIOPNLNSNSSELNSESOSAKAA----- 588
DB 677 -VLPYNTIVREBSLKFNPYDSSRRQOMADMAKIKLSVNLSEBELNRMOSQAKQDAGR 735
QY 589 -----DDPENGERSHTPVSIQEIIVGDFTSKSTGE 620
DB 736 VSEFNSIIRDSLRRKRSVDLEPGEVPSDSEDEGKSHSP-----RASA 780
QY 621 LSESPAGAGAGSTRILIRLNPDSKLSQCVAAAHANANKLFKCG-KEVLVNSO 679
DB 781 LVE-----SSRLSLDLRDREKDLRE-RDERLSSSL-ERNKFYSFALDKTIIPDK 828
QY 680 GEISR---LSTKKEVIMKGNINNYFKLGQEGKYRVYHNOY----- 716
DB 829 ALLERKAKSLSSRE-----ENKSLTMDSDRPAFRNNKDKKEKXDSAPRPIPSWTMKK 881
QY 717 -----STNSFALNKQHREDHDKRHL-AHKFCLTPAGEFKWNGSVHGSKVLITSLRLT 770
DB 882 KIRTDSEGMDDCKEDHKDEEQEROLELFASTR-----LH-SSEIFQDSKRL- 926

```

```

QY 771 ITQENNIPISSFF-----HPNMASHRANWIKAVOMCSKPREPALALALILECAVKNVMLP 825
DB 927 -QHLEKREEDSDPISGLIYKQISEGAN-----STDSIOEPVLLF- 966
QY 826 IMREFLIGHTRLHMTSIEREKEKVKKEKKOBEETMQOATVKKYFPVKHVMKQKE 885
DB 967 -----HSRFMELTRMQOKEKQPKPEVEKQEDTENHP-----KTPESAPENKOS 1012
QY 886 EYRV---TGYGWSMISKTIVYRFVRLPGNTVNRKSLSEG---TKNNDEMMDSDKR 939
DB 1013 ELKTPPSVGPSTVTVTLSEAPSALEKTTGD-----KTVADLVTEKTEPATVSEBA 1066
QY 940 KCSRSFKIKIEPDSEKDEKVGSDAAKA-----DQ----- 970
DB 1067 KPASEPAPAVEDLEQVDLPPGADPPKEAAMMAGVEEGSSGQPPYLLAKPPTPGASPS 1126
QY 971 -----NEMDISKITEKKDQVYKELLDSDSKPCKEEMVEDDMKTESHVNCQ---E 1019
DB 1127 QAESNVDPPEPDSYQPLSKPAOKSEENEPKAEKPDATADAPDANOKAAEAEPSCPAGE 1186
QY 1020 SSGQDVVNVSEGFHLRTSYKKT-----KSKLDGLLERRIKQFTLE-- 1061
DB 1187 DLEVDPVPAKDKPKPKSKRSKTPVQAAAVSTYKAEVTRKSEKID--REKLRSNSPRG 1243
QY 1062 EKQRLERIKLEGGIKIGKITSNKSNLSESPVYTRAKKEGCSDSMKROSPMANNQPE 1121
DB 1244 EAQKLELKEWAE--EKITRASKNSAADLEHPEPSLPLSTRRRNRNVSYATIMGHENNS 1301
QY 1122 DLIOGQSQSDSSVLRMSDSHTINKLYPKDVLVDVSI-----RSPETKCPKONSIE 1173
DB 1302 PVKEPVEQ-----PRVTRKRL--ERELOEAAVPTTPRRGRPKPTRRRADEEBE 1348
QY 1174 NDIEEKVSDL--ASRGOEPLFKST-----KG-NDFFIDSKLASADDIGTLICNNK 1221
DB 1349 NEAKPEAEITLKPEGRKRSRSQKTAAGGPOGKKGNKNEKVADITREAITTEGPOIGVKE 1408
QY 1222 KPLIOESDTIVSSKSALHSSVPKSTNDRDATPLSRAMDFGKLGDCSESNSTLENSD 1281
DB 1409 SSMERPAAEAEAGSEQKRDRIKDPETAPVAVV--EKKPAPEKNSKSKGRSRN 1465
QY 1282 TVSIODSSEEDIMVONSNSIS-----EOPRTREDQVLEPLKCELVGSESTGCEDRL 1336
DB 1466 SRLAVDKSAS--LKNVDAVSPRGAAAGAESEGVAVSPKES----- 1508
QY 1337 PVKGTENAKKPSQOKLEERPVNKCSDQIKLNTTDKKNNNRSESEKKGQSTFQJNG 1396
DB 1509 -----SPQKE-----DGLSSQLKSDPVDPDKEPEKEDVYASGPPPATQL-- 1548
QY 1397 KDNKPKIYIKGECLAKETISRVVSGNVEPKVANNIKIIPENDIK-----SLTVKESAIRP 1451
DB 1549 ---AKOMLEQAVEHIAKLAEASASAAKADAPGIAPEDRKPAHQASETELAAIGS 1604
QY 1452 FINGQVIMEDFN-----ERNSEFKSHLSSSDAG 1482
DB 1605 IIN-DISGEPEFPAPRPYPPGSOITDLOPPAGAQLQSEBEMETDEAVSGILEEATE 1663
QY 1483 NYRDSLETL-PS-----TKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVKKATSSPI 1536
DB 1664 SSRPVPYNAVDPAAGPTDTEANGNSSET--SHSVPEAKSGKEVEYLVKDKROKRTT--- 1719
QY 1537 TSEESNLSNDFIDENGLEPINKNEVNGESKRRKITYITEVTTSTIVATESKTIVIKVERGD 1596
DB 1720 RSRRRKNNKKVAVAVESHVSPESNOAGES---PANEGTIVQHEADQOE-----EKOS 1770
QY 1597 KOTVAVSMENCAKSTVTTTITVTKLTSPSPGSDIISYKROKSTVTTVTVTSLLTTG 1656
DB 1771 EKPHSTPPQSC-----TSDLSKI--PSTENSSQEIETSVBERPT--KASVPDLP-- 1817
QY 1657 GTLVATSMVSKESYSTDRKVKLKMFSPKKTSGTALPYSYRKFTVSTKSKSIFVLJPNDDLK 1716
DB 1818 ---PQPAVDEBPQARFVYHITIESDPVTPPSDPSIP-----IPTLPSYIAA 1861

```



```
QY 1717 KLA---RRGIREVYENYNNAKPAIDIMYPSRPRTFGITMRYRLQTYKSLAGVSLMLRL 1773
Db 1862 KLSPPVASGGI-----PHOSP-PTKYTEWITRROEPPAOSTPS----- 1898
QY 1774 LMASLRMDMAKVPBGSGSTR-----TETSETELTTEILKRDRVGYGIRFEYCIRK 1827
Db 1899 --PALPPTKASDVOTSSSTLRLKIMDKRYASATVSTISV-----TTA 1940
QY 1828 IICPIGVEETPEKPTPQKRLRSSALRKRREPTKQGTGPVLIETWVAEELELWEIRAF 1887
Db 1941 IAEVSAAPCLHEADPP-----PVDSKRLPEKTAAPV-----TNNSELOASEVLVA 1987
QY 1888 AERVKEKAQAV-----EQAKKRLLEQKPPVIAIISTISPTSTST 1929
Db 1988 A---DKEKVAPIAKRITSVRMFVSIDLENSOKITLAKPAPOQLTGLVASLTGLVNS 2044
QY 1930 ISPAQKVAWAPISGVTGTGKVLTLTKVGSPPATVTFQONKNHQEFATWVKGQGN--SG 1987
Db 2045 LVPV-NALKGPKVGSVTT-LKSLVSTPRAG-PVNVL-----KGPVNVLTG 2085
QY 1988 WVQV---QOKVLGIPSTSTGTSTQOTFTSOPRTATVIRPN--TSGSGTTSNSQVIT- 2040
Db 2086 PVNVLTTPVNAVGTVMNAGTVNNAASAVNATASAVITVTAAGAVTAASGCVATITGVIM 2145
QY 2041 -GPOIRPGMTYIRTPLOQSTGKALIRTPVMYQGAPOQVMTQIIRGQPVSTAASAPNTV 2099
Db 2146 AGAVTAPS---TKCKQ---KASANENSREFHGS---MPVIDDRPADAGSAGLRY 2191
QY 2100 SSTPQKSLTATSTSTNSOSSAOPRPOQGOVKLTMALQTLQJHGNOCGLTVIQO 2159
Db 2192 NTSEVNLIS-----YSQKTEGFOR---ISAKISQIPRA---SAMDEIQ-- 2231
QY 2160 GOTTGOLILIPQGVTVL--PGPGQOLMOAAMPNGTVOREFLPPLATTTATTAATTTT--- 2213
Db 2232 -QSVKSKQVPRPSVYASQPSKGPQ---APAGYANVATHSTVLTLQOTYASVYISSV 2285
QY 2214 -----TVST--TAAGTGE-OROSKLSPOMQVHODKTLPPA---OSSSVG 2251
Db 2286 KADRPSEKPEPIHLSVSTSPYOGGTGVYVLTQGINTPRVLYHNLQVLTPLSIYTTNKKLAD 2345
QY 2282 PA--KAOPOTAGPSARPQOPQOPSPAP-----EVQOTPEVQOTQTVSSHPSSEQ 2301
Db 2346 PVTLKIEKVLQAPNAGSTLTPHNPALPSKLTPTVNHHPSSPS:PADRTV-SHL--AAA 2402
QY 2302 PTHAOSKRPVAAOQPOSNMVQSGSPRVQSPQTRIRPSTPSQSLSPGOQSOVQTTTSP 2361
Db 2403 KLDAHSPPR-----SGRPSPPRAS---HPSSTASALSTNATVMLAAGTP 2446
QY 2362 IP-----IQPHTSLOIPSGOPQ-----SQPOVOSTOTLIS--GQTLNOVSVSSPSRP 2408
Db 2447 VPOFISIHPEQSVIMPHSITQTVLSHLSQGEVEMNTPLPSITYSIRPEALHSRAP 2506
QY 2409 QLOQIOQ-----POPOVIANPOLOQ-----VOVLSQIOSQVY---A 2441
Db 2507 -LQPOQIVVARQORASTQOPAPAGPALASQHPREEVHYHLPVARADAPVQSEVLVMO 2565
QY 2442 QIOAQOSGVPOIKLQPLQIOIOOSSAVOHTQIONVTVVOAASVQO 2487
Db 2566 EYRLHPYVPRDVIRIMVPHVTAVSE-QPRADGVYKVPSPASKAQO 2610

RESULT 6
US-09-738-973-425
; Sequence 425, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Radoch
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
```

```
; APPLICANT: Inditias, Carol Joseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliott, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738.973
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 425
; LENGTH: 4019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-425

Query Match 2.7%, Score 402; DB 10; Length 4019;
Best Local Similarity 17.3%; Pred. No. 8.8e-10;
Matches 510; Conservative 314; Mismatches 992; Indels 1126; Gaps 108;

QY 254 CAVCHKL---GDLICETCSANYHLECYKPLIEVPEDEMOCEVCAHKVPVTCVAEI 310
Db 65 CACGKATDPGRLLLCDODISYHYCLDPLOTQVPGKGCWKWCRCNGCATAGLRC 124
QY 311 QNKKPYIRHEPIGYDRSRKKYFWLNRLIEEDTENENKKIMYVSTKQVLAELIDCLK 370
Db 125 EMQNNYTQCAPCA-SLSSCPVCYRNYR---EEDL-----LIQCRC 161
QY 371 DYWEALCKLIEEMBEELIHRMDITEDLTNKAAGSKSFLAANEELIESIRAKKGDIDN 430
Db 162 DRWMAVACQNL-NTEBEVENVADIGFDCS---MCRPYMPAS-----N 199
QY 431 VASPEETEDKNETENDSKDAKNEEPEEDOSLEKSDSKTPDDDPBQSKSEPELVGDK 490
Db 200 VPSDCCB-----SSLVAQIYTKVELDPKRYTQDGV----- 232
QY 491 GNSVANLGDNTNATSEETSPSEGRSPVGCSETFPPDSNMAE-KVASSELPQDVPPEPN 549
Db 233 -----CLTE-----SGMTQLOSLVTVPRRRKRPRK 258
QY 550 KTCESN-TSATTTISIQPLENSNS--SSELNSSQSESAKAADDPENGERSHTPVSIQEE 607
Db 259 IKLKLIINQSVAVLQTPPDIOGSEHSDRGEMDSRGELMDCD---GKSES---SPERE 310
QY 608 IYGDPTSEKSTELSESGACGAGSSTRITRLNPNPSKLSQLKSQVAAAHAKNLF 667
Db 311 AYVDET-----KGVGTDGVKKRKRKP-----Y 333
QY 668 KKGKVLVNSQGEISRLSTKKEVIMK---GNINNYFKLOGEKYRYVHNQYSTNSPALN 724
Db 334 RFGIGFVNRQSRKSGQGTAKRSVIRKSSGSIISQLPCRDG----- 376
QY 725 KQHREDDHKRHLNAKCELPAGEFKWNGSVHGSVLTISTRLITLTOLENNIPSSFFH 784
Db 377 ----- 376
QY 785 PNWASHRANWIKAVOMCSKPREFALALAILCECAVPVYVLMPIWREFLGHTRLHRTSTER 844
Db 377 -----WSQOLDPTLVDESVSV-T 393
QY 845 EEKEKVKKKKKOE--EET---MOQATWVKYTFPVKHOVKKQGEYRYTGVGSGWMI 898
Db 394 ESTEKIKRKRKKKLEETPPAYIOEAFPGKDLADTJRQ-----SKI 436
QY 899 SKTHYREFVPLPGNTNVNYSKLEGTNNMDENDESDDRKCSRPKIRIEPDSERDE 958
Db 437 SLDNLSDEGADQLTYTNNN-----TGLDPSLD--PLSSSSAAPRK----- 475
QY 959 VKGSDAAKGAQONEMDISKITEKKDQYKELIDSDSPCKPEEPR-EVDDMKTESHYNC 1017
Db 476 ---SGTHGPADDPADLADISEVL-NTDDIILGIIISDLAKSVSHSDIGVPTDPPSSLPQPNV 531
```


QY	1018	QESSQVAVVAVWSGPHLRTSYKKKTKSSKLLGILBERIKQFLFLEKOKLEKIKLEGGIKG	1077
Db	532	MOSS-----RPLSEQDLGL-----SPDLKMTVDALLG	562
QY	1078	-IGRTSTNSKRNLSSEPHITKAEGCOSMSRQESP-----NANNDOPEDLIOGCSQ	1129
Db	563	KLYXITPELGGKDVDEFTYAVLSPANQPIPLPQPPPPQLLTHHODAFSRHPLMNGIG	622
QY	1130	SDSSVLEMSDP-----SHTTNKLYPKDRLVDVYSISPEPTKCPKON-----SIEND	1175
Db	623	SSPHLPNHSJPPSGGLGTFESAIAQSSYPAR--DKNSAFNPASDPNMSWTSAPTYGE	680
QY	1176	IEEKVSDLAGRGEPTKSKTKGNDFFIDSKLASADITLCKN--KKPLIOEEDTIVS	1234
Db	681	-----NDTMSNAQRSLTKMK-----EALGEMATYAPVLYTINPINKLEEPPDMWT	728
QY	1235	SSKSGALHSVPKSTNDDEATPLSRMDFEGKIGCD--SESNSTLENSSDVYSIODSE-	1290
Db	729	RVKQ-IAKLWRKASQCRAPRYQYKADNRAALRINKVQSNDSMKRQOQOOSIDPSRID	787
QY	1291	-----EDMIYQNSNESISEQRTREQ-----DVEVLEPLKCE-----	1322
Db	788	SELFKDPDKRESH-EQEKKFRQOKRQSKQOAKIETATOKLEOVKNQOQOQOQOFGSQ	846
QY	1323	---LVSESTGNCEDRLPYVGTGANK-KPSQO-----KLEIRPVNKCSDQIKLNT	1371
Db	847	HLVQSSQDTPSSISQPLPPOPGNGMSPADSFKHELTFRKPPSTPTSTSSDVFVAPQ	906
QY	1372	T-----DKKNNESESEKKGQRTSTFOINGKDKP-KIYUK-----	1406
Db	907	APPPAPSRPIQDSSIQOTQSPSPQVFSGSSNSRPPSPMDPYAKMGTPPRPVYG	966
QY	1407	-----GGLKEISESVVSGNVEPKNNINKIIPENDI--KSLTYKESAIRPF	1452
Db	967	HSFRRNSAAPVENCQPLLVSSRPLDMN-----ETTANRSPVRLCSSSTINDPYAKP	1022
QY	1453	INGDVIMEDFENERSSEFTKSHLLSSDAEG-----NYR	1485
Db	1023	DTPRPVMTQDPKSLGHSRBPVYSBQTANGPIAAGSHFTKPSRADVFQOKRIPDSYA	1082
QY	1486	DSLETLPSYTESDSTQTTTYSASCPES---NSVNOV-EDMEIETSEVKVYTSPTISEE	1540
Db	1083	RPLLT-PAPIDSGCPKPTQMOPPPSSODPYGVSQASRLLEVDERALPRPINDFS	1141
QY	1541	ESNLSNPFIDENGPIPKKNENVAGSEKRTVITEVYTMNSTVATESKYVIKXENKDKQY	1600
Db	1142	H-NOSNDPYQSP--PLTPHPAVN-----ESFAPHSRAFSOPGTISRPSODPYSQ	1188
QY	1601	VVSST-----ENCAKSTVTTTTTIVTKLSTPSTGSDVILISVASESKTVTTVTTLSITTY	1655
Db	1189	PGGPRPVVDVYSQSSOSTANSNMDPYSQPGPRPRPTTVDPYSQQTGRPSIQDLPFTR	1248
QY	1656	GGTLVTSMTVSKYSTRDYKVKLKKESRPKTKNSGALPSRYKFTVKSTKSLIEVLPLNDL	1715
Db	1249	---VTNQRHSDPYAHPGTPRGISVPSQAPATPRPRISGFTKNSMTRPVLMPND-	1303
QY	1716	KKLARKGIEVPEFN--YNAKALDIWYPPSRPRTFGITMKYRLQTYKSLAGVSLMRL	1773
Db	1304	-----PFLQAAQNRGAL--PGQLVPR-----	1323
QY	1774	LMASLRWDDMAAKVPPGSGSTRTSETETITTEILIKRDVQPGYIFREYCIRKTIICPG	1833
Db	1324	-----	1323
QY	1834	VPEPKETPTPQKGLRS--SALRPKRPTPKQGTGVILIEWAAEELEMEIRAFARV	1891
Db	1324	-PDTCSTQPPPPGGLSDYFSRVSPEAADPYQDSMPTRS-----	1363
QY	1892	EKEKAQAVEQAKRLEQCKP---TYIARSTISPTISSTISPSAQKVMVAPISSVTYG	1948
Db	1364	-QSDSFTQSTAHDAVADQPRPGSEGSFCASSNSPMHSQGOQFSYSQ-LPGVPTSGVTD	1421

Dd	773	QAQ-----IDQAPLMPD-----	TTNEEVAEALERINAAKVSVKAIETAITTAQDLERVK	821
Qy	1340	GTE-----ANGKKPSQOQKLEERPVNKCSDOIKLNTTDDKKNNEERESEKGGORTSTFOIN		1395
Dd	822	NEEISIKENIENDSTQTKMDAYNEVKQAAATARKAQNAT-VSNATNEEVADEADAQAQO		880
Qy	1396	GKDMPKLYLGEGLCELTSESRSVSGVEKPEKVNIN--KIIPENIKSLYKESAIRPF		1455
Dd	881	GLHDIQVQVKS-----QEVADPK-----SKVLDKINAIDTOAKVPAAD---TEVENA---		925
Qy	1453	INGDIVIMEDFERNRSSETKSHLSSDAEGNRYDSLETLPSTKESDSTQ-----TTTPSAS		1508
Dd	926	-----YTRKROQIÖNSNASTSEKQAAYTE-----LDTKQÖARTMLDAAINNDSVT		972
Qy	1509	CPENS---VQVQEDMELETFSEKKYTSPTTSEESNLSNDEIDENGLIPKNENNVGE		1566
Dd	973	TAKDMSIAINQVOAATTKKSDAKA-----EIAOKA		1003
Qy	1566	SKRKIVIEVTTMTSJTVAETSKYIVIKEDKQOTVASTE-----NCAKSTVTTTT		1618
Dd	1004	SERKAIETAMDST-----TEEQQAK-DKDYQAVYANMADIDNAAANNVDNAKTTNEAT		1055
Qy	1619	VTKISTPSTGSGVDIISVKEQSKVTVTVTYDLSLTTGGTGLVTSMTSKSEYSTRDVKYLM		1678
Dd	1059	IAAI-TPDA-----VYKPAKQALIDKQVOEETIDG-----NNGSTEEKAAK		1102
Qy	1679	KESRPKTKRSGTAL-PSYRKPFVTKSTFKSLFFVLPNDDLKIKARKGIREVPYFNNAKPA		1737
Dd	1103	QOVQEKTKTAAALDAAHNAEVEAAKA-----AIAIEAIOPATTTKONAEYA		1155
Qy	1738	LDIWPSRPRTFGITWRKYLQTVKSLAGVSLMLRLMLASLRMDMAKVPGGGSTRTE		1797
Dd	1153	I-----ATKAN		1158
Qy	1798	TSEETITTELIRKRDVGPGRIFREYCIKRIIOPIGVPEPKETPIPOKKGLRSSALRPK		1857
Dd	1159	EKKTAIOQTUITLEETLAANADVDNAVTOANSINPAANSÖNDVDAKTTGNSI-----		1213
Qy	1858	RPEPKQTPVILITWVAEEL-----ELMEIRAEERVEKEKAQVQEQAKKRLLEQO		1911
Dd	1214	-----DQVTPVNNKATARNBEITAILNNKLOEIOQANPPATDEEK-QAADAENTENGKAN		1267
Qy	1912	PTVATSTSTSTSTSTISPA-----QKMAVPJSGSVTTGKXVLTITKVGSPATVFO		1966
Dd	1268	QAI SATTNAAOVDEAKNAAEAPAINAVTPPKVKKQQAKKDEIDLOAQOTQVWINNDONATTE		1327
Qy	1967	ÖKNNEHQFETAVKQOGSN-----SGVVOVQKVLIGLIPSTGSGOOTFSPQPRAT		2019
Dd	1328	EKEAIIQOLAAVAIDAKNNITTAATDDNGVDQAD-----AGKSIDISTOPATVAV		1376
Qy	2020	VTRPNTSGSGGTSNSOV---ITGPOIRPGMVIKRTPLQÖSTFLGKAIITRPVWOPGAPQ		2077
Dd	1377	KSNAKNDVDAQVTTQONOAINTTG-----ATTEEKAANKADLV---LKAKEKAYQ		1422
Qy	2078	QVM-----TOIIRQOPS---TAVSAPRTVSSTPGQKSLTSATSTISIGSASQAPRP		2127
Dd	1423	DILNAQÖTTNDVTO-IKQÖAVADIOGTIADPTTINDAKDELAKRABQKALLAQTQDAFTE		1481
Qy	2128	QÖQGVKLTLM-AOLTOIOLPGHGQMGVLTVYIOGQGTQÖLOLIPQGVVLPBPGQÖOLQA		2186
Dd	1482	EKEQANQOVDAQOLQÖGNQÖNIENQÖSIDDVATKADNAIQADIDPLQÖASTDVKTAARELLT-		1540
Qy	2187	AMPNGVQÖRELPFLPATTAFT-----ASTITTYVSTTAAGTG		2223
Dd	1541	-----ENQÖKITTELINNNETNEKGNDIGPVAAYEELINNAINATTTGDTV-TYAKDTA		1594
Qy	2224	EÖQÖSKSLSPQÖQVQÖDKTLPBPQÖSSVGPAKAPQÖPQÖASBPQÖFQÖTQÖSAPQÖEVYTO		2283
Dd	1595	VQK-----VQÖLHANVYKRPAGKKEIDQÖAADPKTI-QÖEÖTPNNSQÖEINDAKEVÖTE		1647
Qy	2284	PEVQÖTÖTVSSHVSEAPÖPTHAQÖSKPQVAAQÖSQPQVQÖSGPYRVQÖSPQÖTRI-----		2338

```

Db 1648 LN-QAKTNVDOSSINEYVDNAVEKGRKAKIA-VKTSSEKKKALAKIEDIYANAKNEADN 1703
QY 2339 -RPSTPSOLSFQGOQYV-----QTTISOPITPIOPHTSL-QIPSGQPOQOPQVQ 2385
Db 1706 SNASTSSIEIAKOKIAELKQATADQNVNATSKDIDIEVOIHNDLINDINDYTIPTGKKESA 1765
QY 2386 SSTQTLSGQGLNVSNVSSPSRPOLOIQOPQPVIVAPVLOQOQYV-----L 2432
Db 1766 TTDLATYADQKKNIISDIT-NATODEKQOQ-----AIKQDVAVQTALESINNGVNDGV 1818
QY 2433 SOIOSVVAQIOIAQOSGVPQOIKLOLPIPIOQSSAVOTHQIONVTVYCAASVOEOLQHVQ 2492
Db 1819 DDALTOGKAALDIAQYDAFVAPKPAQAIEVKADETKESIDQSDQJLAEKTT--FALAMIK 1876
QY 2493 QLRPOQ-----QKKQOQOIEINVTTPSKLLIKVELIQOYMKHNAVIEHLK 2539
Db 1877 QITDOAKOGITDATTAETAEVEKARAQGLEAFDN-----IJDISTEKQ-----KATIELE 1924
QY 2540 QKKSMTPAEREENORMIVCNOVMKYLIDKID-----KEEKQA-----AKRKR 2582
Db 1925 T-----ALDQIEAGVNNNADATTEBEKAFTNALEDLSAT 1960
QY 2583 EESVEQKRSKONAT-KISALLFKHKQOLRAELI-----KKRAL-----LDKDLQIEVOE 2631
Db 1961 EDISDQTTNEIATVKNASAL-----BOLKQORINPEVKRKALEAIREVYNNK--QIEIIRN 2013
QY 2632 LKRDLKIK-----KEKDLMQADATVAAPCPVPIVAPAPAPSPSPPPG 2679
Db 2014 ADADASKELIARTDGLGTFPRPADKIDKQTNLEVALEQNV--IPALAIYPOQNDPAN 2071
QY 2680 VOHTGLISTPPLVYASQKRREBEKSSSKKKKKKIISTSEYTKKDT 2727
Db 2072 DTNNGI-----DNNPATANSANANATPENTGQPNVSEI 2103

```

```

1 RESULT 11
2 US-09-815-242-12967
3 ? Sequence 12967, Application US/09815242
4 ? Patent No. US20020081569A1
5 ? GENERAL INFORMATION:
6 ? APPLICANT: Haselbeck, Robert
7 ? APPLICANT: Ohlsen, Karl L.
8 ? APPLICANT: Zyskind, Judith W.
9 ? APPLICANT: Wall, Daniel
10 ? APPLICANT: Trawick, John D.
11 ? APPLICANT: Grant J.
12 ? APPLICANT: Yamamoto, Robert T.
13 ? APPLICANT: Xu, H. Howard
14 ? TITLE OF INVENTION: Identification of Essential Genes in
15 ? TITLE OF INVENTION: Prokaryotes
16 ? FILE REFERENCE: ELTRA.011A
17 ? CURRENT APPLICATION NUMBER: US/09/815,242
18 ? CURRENT FILING DATE: 2001-03-21
19 ? PRIOR APPLICATION NUMBER: 60/191,078
20 ? PRIOR FILING DATE: 2000-03-21
21 ? PRIOR APPLICATION NUMBER: 60/206,848
22 ? PRIOR FILING DATE: 2000-05-23
23 ? PRIOR APPLICATION NUMBER: 60/207,727
24 ? PRIOR FILING DATE: 2000-05-26
25 ? PRIOR APPLICATION NUMBER: 60/242,578
26 ? PRIOR FILING DATE: 2000-10-23
27 ? PRIOR APPLICATION NUMBER: 60/253,625
28 ? PRIOR FILING DATE: 2000-11-27
29 ? PRIOR APPLICATION NUMBER: 60/257,931
30 ? PRIOR FILING DATE: 2000-12-22
31 ? PRIOR APPLICATION NUMBER: 60/269,308
32 ? PRIOR FILING DATE: 2001-02-16
33 ? NUMBER OF SEQ. ID NOS: 14110
34 ? SOFTWARE: PasteSeq for Windows Version 4.0
35 ? SEQ ID NO 12967
36 ? LENGTH: 2478
37 ? TYPE: prt
38 ? ORGANISM: Staphylococcus aureus

```


QY 2386 SSTQTLSGOTLNVGSVSSPSRPOLOIQOPQVIAVPOLQOVY-----L 2432
DB 1766 TITLYAYADOKNNNISADT-NATODEKOQ-----AIKOVQVQVIALESINNNGVNDY 1818
QY 2433 SOIQOVVAVOIAQOAGVPOQIKLOLP1QOQSSAVQTHQVNVVVOAASVOEOLQRY 2492
DB 1819 DDAITGOKAIDAIQVDAVYKPRANQAIKADTKESIDQSDQLAEEXT--EALAMIK 1876
QY 2493 QLFDOQ-----QKKQOQEIINVNPSSKLIKVELIQOVYKNAVIEHLK 2539
DB 1877 QITDQAKQGITDATTAEVAKAAGLEAFDN-----IQIDSTEQO-----KALIEELE 1924
QY 2540 QKSMTPAEREENQRMIVCNQVAKYLLDKID-----KEEKOQ-----AKKKR 2582
DB 1925 T-----ALDQTEAGVNVNAADTPEKRAFETALDILSKAT 1960
QY 2583 ESEVEQKRSKONAT-KLSALLFKHKEQLRAELI-----KKRAL-----IDKLOIEVOEE 2631
DB 1961 EDISDQTTNAETVKNASAL-----EQLKQRIINPEYKKNALEIREVYVK--QIEIILKN 2013
QY 2632 LKRDLLIK-----KEKDLMOQATAVAPRCPVTVLAPRAPPSPPPRG 2679
DB 2014 ADADASAKETARTDLGRYEDRFADKLDKTQTNAEVLELQNT--IPATEIIVQNDPDAN 2071
QY 2680 VQHTGLSTPTLPVASQKRRKEEKDSSSKKKMIISTSKETKKDT 2727
DB 2072 DTNNGI-----DNDATANSNMNATPENGPVNSBT 2103

RESULT 12

US-08-681-219-32

; Sequence 32, Application US/08681219
; Patent No. US20020058607A1

; GENERAL INFORMATION:

; APPLICANT: Takaaki Sato and Junn Yanagisawa

; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN

; SIGNAL-TRANSDUCING PROTEINS AND THE GLGF

; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/681,219

; FILING DATE: 22-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: White, John P

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JKM

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2843 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-681-219-32

Query Match 2.4%; Score 361.5; DB 8; Length 2843;
Best Local Similarity 17.3%; Pred. No. 3.e-08;

Matches 412; Conservative 325; Mismatches 846; Indels 795; Gaps 100;
QY 373 WEAEELCKILEEMEEIHRHDDITEDLTNKRAGSNKSEFLAANEELIESIRAKGDDINVK 432
DB 699 WDMGAVSMKLNLIHSKHKM-----AMGS-----AALRNLMANRPARYKKA-NIM 743
QY 433 SPE-----EPFEKDK---NET-----ENDSKDAENREFEEDQSLKED----- 466
DB 744 SPQSSLSPLHYRKQKALEAELDAQHLETFEDNIDNISPRASHSKQRHKQSLXGDIYVFDI 803
QY 467 --SDDKTPDD-----DPEQKSEEPTEVGDKGNSVANIGDNTNATSEETSPSEG 515
DB 804 NRHDDNRSDNFNTGNMTVLSPYLTNTVLPSSSSRSGLDSSR-----SEKDRSLER 854
QY 516 RSPVGC-----LSETPDSSMAEKRYASELPQ--DVPEPEN--KTCGSSNTATTTISI-- 564
DB 855 ERGIGLGNHPATENPGTSSKRGLOISTAAQAKWEEVSAIHTQOEDRSSSTTELC 914
QY 565 ---QPNLENSSSSELNSSQSESAKADPENGERESHPTVSTOEITVGDFTESEKSTGEL 621
DB 915 VTDERNALRRSSAAHTHSNTYNTKS---ENSNRTCSMPYA--KLEVKRSSNDLSNV 967
QY 622 SESPGAGKAGSGSTRITIRLNDPSKLSOLKS--QQVAAAHEANKL-FKEGKEVLVN- 677
DB 968 SSSDGYGKRGOMKPSIEYSSEDEDEKFCGYQYPADLAHRIHSANHMDDNDGELDTPINY 1027
QY 678 -----SQGEISRLSTKKEVIMKGNITNNYKLOQEGKYRYHYHNOYNSPFLAKHQH 728
DB 1028 SLKYSDBQLNSGRQSPSONERMAPKHITIEDIKOSBORSR--NOSTI--IPVTTEST 1082
QY 729 REDHDK-RRHLAKHFCLTAPAGEFKWNGSVHGSVLTITLRLITOLENNIPSPFFHPMW 787
DB 1083 DDKHLKFGPHFGQGEVSPYRSGANGS-----ETNRVGS----- 1117
QY 788 ASHRAMWIKAYQWCKSPREPRALALIECAVYVWMLPIVREFLGRHLRMTSIREEK 847
DB 1118 -NNGINQVNSQSLQED-----DYED 1138
QY 848 EKVKKKEKQEEETMQOATWVYTFEPVKHQVWKQGEERYVYGWMSIKTHYRYV 907
DB 1139 KPTNYSERYSEEL-----QHDEERPTNY----- 1162
QY 908 KPLPGNTVNYRKSLEGTKNMDENDSE-----DKRCKSPKKIKI 950
DB 1163 -----SIKYNEE---KRVHDPIDYSLIKATDIPSSQKQSFSPKSSQSSKSTEMH 1211
QY 951 EPDSEKDEYVGSAAKGAQDNEMDISKITPEKKODVKELDNS-----DKP 997
DB 1212 SSSSEMTSPSSNAKR--ONOLHPSSAOSRSGOPQAAATCKVSSINQETIQTVCVEDTP 1268
QY 998 -CKEEMPEVDDMKTESHVNCQESSQ-----VDVVNVSEGFHLRT-----SY 1038
DB 1269 ICSRSSLSLSSASDEIQCNOTQOEDANSANTLOIAIEIKIGTSAEDPEYVAVSG 1328
QY 1039 KKKTKSKLDGLERRIKQFTLEKQRLKIEGKIGKIGTSTNSSKN-----LSESP 1093
DB 1329 HPRTKSSRLQG-----SSLSESARHKAVERSGAKSPSKGAGTQPKSPREHYVQETP 1381
QY 1094 VI-TKAKGQSQSMQEQSPNANNQOPEDLIGCQSQSDSVLRMSDPSHTTKLKPXKR 1152
DB 1382 LMFSCRTSVSSLDSPFSRSTASSVQSEP-----CGGWVSGIITSPDLDPSPQOTMPSPR 1435
QY 1153 VLDDVSLRSPETKCPKONSTIENDIEKVSDLASRGQEPYTSKTKGNDFFIDSKLSAD- 1211
DB 1436 -----SKTPPPPPQTQOTKRE-----YKKNAPITAEKESG-----PKQAAVNA 1474
QY 1212 -----DIGTLI-----CKNK-----KPLQOESDITV-----SSSK 1237
DB 1475 AVQRYQVLPDADPILLHPATESTPDGSCSSISALSLDEPFQIKVDELRIIMPVQGDNDG 1534
QY 1238 SALHSSVPKSTNDRDATPLSRAMDFEGKLGQSESNSTLENSDITYIQDSSSEDMATYON 1297
DB 1535 NETESDQPKESNENQ-----EKEAEKTIIDSEKD---LLDDSDDD----- 1570


```
QY 1298 SNEISIEQFRTQEVLEVLKCELVSGESTGNCEDRLPVKGTGTEANGKKPSQOKLEER 1357
Db 1571 -----DIEILE--EC--IISA-----MPTKSSR-KAKKPPAQTASKLPP 1603
QY 1358 PVNKKSDOI---KLNTJDDKNNENRESEKKGQRTSTQINCKDKPKYLKGECLAKES 1414
Db 1604 PVARKPSOLPYVKLLPSONRLOPKHVSFTPG-----DDMPRYV---CV----- 1644
QY 1415 ESRVSGVNEPVKNINIKIIPENDIKSLTVKESAIRPFINGDVIMEDFENRNSSETKSHL 1474
Db 1645 EGTPINFSTATSLDPLTESPPNEL-----AGEGVRGGAQSGEFKRTIPIEGR- 1695
QY 1475 LSSSAEGNRYDSELTLESTAKES-----DSTQTT 1503
Db 1696 -STDAOGGKTSSV--TIPELDONKAEEDILAEICINSAMPKGSKHPFRKKIMQVOQA 1753
QY 1504 TPASACPESNSVNOYEDMEIETSEVKKYTS-----SPITSE 1539
Db 1754 SASSSAPKKNOLDGKK--KKTPSPVKPIPONTREYTRVRKKNADSKNNLNLAERVSDNKS 1811
QY 1540 EBSNLNDPIDENGLPIKNENVNGE----- 1565
Db 1812 KKNLKNNSKNDKLPNNEDRVRSFAFDSPHHTPIEGTPYCFSRNDSLSLDFDDDD 1871
QY 1566 ---SKRKVITEVTMTSTVATESKTVIKVENGDQTVVSTENCASKTVT----- 1613
Db 1872 VDLGREK---AELRAKERESEAKVTSHTELTLSNOQSANKQALAKQINGGQPKPIQ 1928
QY 1614 -TTTTTVTKLSTPSTGSGSDITISVEQSKTVTTVTDSLTGTGGLVTSMTVSEYSTR 1672
Db 1929 KQSTPQOSKIDPDGKAATD---EKLONFAIENTPVCFS--HNSLSLSLSDIDENNNK 1982
QY 1673 DKVKLMKFSRP-----KTRSGTALPSYRK-----FVTKSTKKSIFLVNDL- 1715
Db 1983 ENEPRIKETEPPOSEBPKPOASGAAPKSFHYEDTPVCFSRNSSLSDISEDDLQDC 2042
QY 1716 -----KTLARKGIREVPYFNANP-----ALDIMPYSPRPRTFGIT-- 1753
Db 2043 ISSAMPKKKKPRLKGDNE-----KHSPRNMGILGEDLTLKDIOQBPDEHGLSPDS 2096
QY 1754 ---RRYRLQTVKSLAGVSLMLRLMASLRMDMAKVPBGSGSTRTESEIEITTELI 1809
Db 2097 ENFDKKAIOEGANSIV-----SSLHQAAAACL-----SRQASDSD--STL 2136
QY 1810 KRWDGPIGIREYCIKRIICPIGVETPKETPTQKGLRGSALRPKPPETPKGTGPIV 1869
Db 2137 SLKS-----G-----SLGSPFHLTPDOEEKPFTSNKGR--ILKPGKST- 2175
QY 1870 IETWVAEEB-----LELMEIRAFARVEKEKAQAVEQAAKKRLDQKP- 1912
Db 2176 LETTKIESESKIGKGGKKVYKSLITGVKVSNE-----ISGOMQOPLDANMPSISR 2226
QY 1913 --TVIATSTSTSTSTISPAQKVMAPIGSAVTGTCKMVLITKVSAPATVTOQKN 1970
Db 2227 GRTMHIHIGVNRSSSTSPVSKKGPPLKTPAKSPSEGTATTSBRGAKPSV- 2278
QY 1971 FHQTFATWKGOSNSGVVQVQOKVIGIIPSTGSOQFTSFQRTATVTLIRPNTSGSG 2030
Db 2279 -----KSELSPYARTSOIG-----GSSK-----APRSRSGR 2305
QY 2031 GTTNSQVITGPQIRPGMTVIRTPLOQSTLGAIRTPVMWPG---APQOVMTQING 2086
Db 2306 DSTPS-----RPAQOPLSRPIQSPGRNS-----ISPRNGNISPPKISQLPRT 2348
QY 2087 QPVSTAASAPNTVSTPQOKSLTS-----ATSTNIQSSASQAPRPQOGQVKYLM 2137
Db 2349 SSPSTA---STKSSGSGKMSVTSFGROMSOOLTKQGLSKNASSIPRSESAS- 2399
QY 2138 QLTQTOGHGNGGLTVVIOGOGTTGOLQ-----LIPQGVTLPGGQOOLMAAMPN 2190
Db 2400 GLNOMNNGANKKVELSRMSTSKSSGSESDRSERPVLVROSTFIKEAPSPILRRKLEES 2459
```

```
QY 2191 GIVQRLFTPLATATATTTTIVSTTAAGTGEOROSKLSF-----QMOVHD----- 2239
Db 2460 AS-----FESLSPSSSRPAPSPRSQAQTPV-----LSLSPDMSLSTSHSVQAGW 2504
QY 2240 KTLPPAOSSVAPAKAOPOTAOPASAPQOPOTOPSPAPQEVOTOPVOTQTVSSHVPS 2299
Db 2505 KLLPPLSLTIEYNGRPAKKRIDIASHESPSRLPINSGTWKRE---HSHSSSLP 2561
QY 2300 AOPTHAOSKPOVAOQOSPOSNVQOSPVVQOSPTQIRPSTPSOLS--PGQOSOVQTT 2358
Db 2562 SWMRRTGSSSSILSASSESEKAKSEDEKHVNSISGT--KQSKENQVSAKGWIRKIKENE 2619
QY 2359 SOPITPHTSLQIPSOQOPQOSPOVOYSTOTLSGOT 2396
Db 2620 FSP-----TNSTQTVSSGAT 2635
```

```
RESULT 13
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSHOT, Astrid Adriana Anna Maria
; APPLICANT: KOHN, Jennifer Leigh
; APPLICANT: WEISS, Bertiam
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 472505
; CURRENT APPLICATION NUMBER: US/09/764,176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the
; OTHER INFORMATION: Ictc acid sequence of AAP-
US-09-764-176-7

Query Match          2.4%; Score 354.5; DB 10; Length 1400;
Best Local Similarity 18.2%; Pred. No. 3.1e-08;
Matches 290; Conservative 240; Mismatches 584; Indels 481; Gaps 56;

QY 274 HLECVKPLIEVPEDEW-----QCEVYAHKVPGTDCVALEIQKKPIRIHHPGY 324
Db 61 HUKMKKIGKSVYADRWKEYLIKYLCEQFDNKLFKNTIN-----EEDADTMRLOPIGR 115
QY 325 DSRRRKYWFL-----NRLLIEDFTENENEKRIWYSTKVQLAELIDCDKDYWEALCK 379
Db 116 DKDGLMYWQLODDHNVRYIEQDDQ----- 143
QY 380 ILEMKREIIRHMDITEDLTNKRGSNKSFLAANEELISIRAKGIDIDNVKSPEBTEK 439
Db 144 -----GSSWKCIVRNRELAETLALLKAOIDPVLKNSSQO 179
QY 440 DKNETENDSKDAEKNRBEFEDQSLKSDSDKTPDDDPQEGKSEEPTEPVDDKNSVSNLNG 499
Db 180 DMSRSPSLDEETKKEETPKQOEOKES-----EKMKSEE--QPMDLERNSTANVL 230
QY 500 DNTVATSEETSPSGRSVGCLETSPDSSMAEKKVASELPQDVPEEPNKTCESSNTSA 559
Db 231 EETTVKKEK-----DEKELVKLPVLYKLEKPLPENENKRIIEEBS 273
QY 560 TTTSIQPLNLSNSSSELNQSSESAKAADPENGERESHHPVSIQIEIIVGDTFSEKSTG 619
Db 274 FKENVKP-----IKVEVKECBADPKDTKSSMEKPVAAQEPERIEFGNIKSSH 320
QY 620 ELSESPGACKGASGSTRITTRLRNPDSKLSQKSOOVAAMAAHAKKLPREGKEVLYVNSQ 679
Db 321 ETTER-----STEETKRLKNDQAKIPLKKREIKLSDDFDSPV--KGPLCKSVYPT 369
```


QY	660	GEISRLSKREKVLIMKGNINNYFLGOEGKRVYHNOYUINSNALKKHQHREHDKRRLLA	733
Db	370	KEFLKDEIKOEELTKCRKISTTLTALGHGK-QLVNNGEVSERYAPN-----FTPEIE	420
QY	740	HKFCULTPAGE-----KNGSVHSGKVLTV--ISTRLTITOTLENNIPSSFPFN	786
Db	421	TKFYETKEESYPSKDRNITTEGNGEINSVITSMKTELEKTAFLPKADSS-----	475
QY	787	WASHRAWIKAVOMCSR-----FRERLALALIECAVKRYVWMLPIWREFLHTRLHR	838
Db	476	-----ISVLEHSHKAOIEBDDPEMETSLDSSEMA-----KDLSSKTAJSS	517
QY	839	MTS---IEREEKYKKKKKOEEOEETMOQATWVYTF-----PYKHOWKOGGEY	887
Db	518	TESTCMGEEKSPKTKDKRPPILCELEKLEKSKTFLLDKOQRSLPIEVPKSTLESE	577
QY	888	RVYITGGMSWISKTHVRYFVPLKPLGNTV-----NYRKSLEGTKN	927
Db	578	KPGSPEAETSPENIIDHCEKELASEREVEECOSTYVGOGVKKVYDLETLKEDESEFKV	637
QY	928	NMDNMEMESD-----KRCSSRPKKIKIETPSEKDEKVGSDPAKGAONEMDJK	977
Db	638	EMD-NLDNAQTSGLIEPSEFKSGMOKSKRYKALVPEEIT-----ASEN---TE	682
QY	978	IT-EKKDQDYKELLSDSDKPCKEEPMEYDDDKTESHYNCOSSQVDVYVSEGFHL	1036
Db	683	ITSEROKEGIKLITRISRRKKKDPSPKV--LEPE--NKOEKTEKEEKTNGVTRLR	736
QY	1037	SYKKKTKSSKLDGLER-----IKOFTLEKORLEKI	1065
Db	737	SPRLSRPTAKVAELRQDKADKKRGEDEVEESTALQKTDKKEILUKSEKTNISKV	796
QY	1070	KLEGGIKIG-----KTSJNSKNLSESPVITKAKECQSDSMROESPANNDDP-	1120
Db	797	KPKKQVWMTGSRTGRKRWKYSNDESGSGSEKSSAASE--EEEKESSEALILADDEPK	854
QY	1121	EDLIOGCSQSDSSVILMSPSHTKLTP-----K	1150
Db	855	KGGLPNHPELILCDSQSY-----FTACLPRMLIIPDEWFPCPCOHLKCEKLE	907
QY	1151	DVRLD-VPSLRSPETKCPKON-----STENDIEKVSIDLASROEPTK--SKTKGN-	1198
Db	908	EQLDODDVALKKKERARERKRELRYVYGISTENIIPQEDPSEDEBEKKKDSKKNALL	967
QY	1199	-----DEFIDDS--KLASAD-----DIGTLCKKKRP---LI	1225
Db	968	ERRSTRTRKICISYRFDEFDEALDEALIEDDIKEKDDGGVGRKDISTIGHROKDISTILD	10277
QY	1226	QESDITVSSSKSALSHSVKPSYNDNDARPLSAMDPEKGLGDSSENSTLENSDTYSI	1285
Db	1028	EERKENKRPORAAARAKKRRRLND-----LDSDSNDEESEDDEFKI	1070
QY	1286	QDSSEEDMIYONSNESISEQFTRTREDVAVLEPLKCELYGSESTONCDBRLPYVKTEANG	1345
Db	1071	SDGSQDEFYVVSDENDPESEDPSND-----SDTFCGRLL--RRHP	1111
QY	1346	KKPSQOKKLEER--PVAKGCDQIKLNTDKKNNR-RESF-----KKG	1386
Db	1112	SRPQRSRRRLRRTPKPKKYS-----DDEEESSEKSRSDSDFDSDFVETRRRS	1166
QY	1387	QRTSTPQINGK-----DNKPRITYLKGCEKLEISESVVSGVGNPYVNNINKITIPENDKS	1441
Db	1167	RRNQKRINYNKDESEDSGSKSLRKEKELIRVHKRRLLSSSESESYLSKNS---EDDFLA	1223
QY	1442	LYVESAIREFPINGDVIMEDFERNSSSEFKSHLLSSDAEGNYRDSLETLPTSKESDTQ	1501
Db	1224	KESKRSYRKKRGRTSTDESEADEEEEEEKPS-----RKRLHRI-ETDEEESCD	1271
QY	1502	TTTTPSACSPSSVNVQYEDMELTSEYKAVYSSP--IYSEESNLNDFIDENGLPIKNN	1559
Db	1272	NAHQDAQPRDSQPRVLSEDS-----TKKPRIESEDEEDEN--VKVSGSPLDYS	1323

```

Oy 1560 ---ENVNSESRRKYITFEVTMTSTVAIAESKTVIAVEGDQOTVSSSENCAKSVTTT 1615
Db 1324 LVDPDSTNQSOPGKAL-----ENLIGKPTKESQTPKDNS 1357
Oy 1616 TTTVTIKLSTPTSGSVSDITISKEQSKVTVTYTTD 1650
Db 1358 TNASLIASNGTSGCGACGAPFEEDDELLVATDLDV 1392

RESULT 14
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match 2.3%; Score 342; DB 10; Length 2368;
Best Local Similarity 17.6%; Pred. No.1,9e-07;
Matches 436; Conservative 342; Mismatches 866; Indels 838; Gaps 111.

Oy 475 DEQGGSEEPTEVGVGKGSVSNALGDNNTNATSEETSPSEGRSPVGCISETPDSSNMAEK 534
Db 40 EGNQPAQNOQAPADANQNPANAGAQ-ANPTAQATPATPANQQPA---AQPANQGOQAMP 95
Oy 535 KYASELPQDVP-----EPPNKTCCSSNTSATTTSTQPULEN-----SNSSSELNSSQSE 583
Db 96 AGGAQOPMTQPPAGQGNQADPNNAQAOPQNOAAPNQAQGNQNOATPNNNANPANOQTOPA 155
Oy 584 SAKKAADDPENGERESHTPVSIOEIVGDPTEKSTGELSESPGAGKAGASGTRITRLRN 643
Db 156 NAPAQAQP-----ANNAQTQDPNASNTEGGSINTLTLPED 190
Oy 644 PDSKLISQKSOQVAAAHEAN--KLFEKGEVLVYNSOGELISRLSTKKEVINKNINNYF 701
Db 191 PAISTDENRQDPVTYVTDKVNQYSLINNGK-IGEVNSE-----LRRSDMFDKNNPQNV 242
Oy 702 KLGGGCKRRVYHNQYSTSFALNKQHQRHEDHDHKKRHLAKFCLTPAGGKKMNGSVHG-SK 760
Db 243 -----QARGNVAAAGRVNANDSTD-----HGFNGNLSK 270

```



```

QY 1810 KRBDVPGIREFYCIKICPI-----GVPEPKETPPQKGLR---- 1850
Db 1167 TAEELIAANANAVQANNNIENANSONDVQAKTGEASIDQVTPVKNKATAVTDA 1226
QY 1851 ----SSALPRKRPPEPKOCPYIETWVAEEELMETAPAFARVKEKAQAVEQAKKR 1906
Db 1227 KNNITPAIDNDGVDAKDKNSIOS-----TOPATAVKSNAKND 1266
QY 1907 LEOQKPTVATSTSTSTSTSPKQVWVAPISVVTGTKMVLTTKVGSPTAVTFQ 1966
Db 1267 VDO-----AVTTQONAIINTGATTEK-----NAKDLVLKAK----- 1300
QY 1967 QKNNEHQTFATWVKOGQNSGVVVOQKVLGII PSSGTSGOQFTSFOPRTATVTIRPNT 2026
Db 1301 -EKAYOD-----ILNAQITNDVTOIK-----DOAVADYQGITADTFIKDVA 1340
QY 2027 SSGGTSNSQYITGPQIRPGMTVIRTPLOOSTLGKAIIRTPVMVOPGAPQOVMTQIRG 2086
Db 1341 KDELAKKAEQ-----KALIAQTADATTEKEEQANQOV--- 1373
QY 2087 QPVSTAVSAPNTVSTPGOKSLTSATSTSNI---OSSASQPPPOQGVKL-TMAQLTOL 2142
Db 1374 -----DAELTQGNQNIENASIDVNTAKDNAIQALDPIQASTDVKTNARABELL 1422
QY 2143 TQGHGNGOGLFVVIQOGQITGOLIPQGVTVLPQPGOQLMQAMPNGTVOREFLTPLA 2202
Db 1423 TE---MOKNITEILINNNETNE---EKNDI--GPVRAVEEGLN----- 1460
QY 2203 TTATTASTTTTSTTAAGTGEOROSKLSPOQVHODKTLPPAOSSSVGPRAKQPTAQP 2262
Db 1461 ---INATTTGV-ITAKDTAVOK-----VOOLHANPVKPKAGKTALDOAAADKKT-QI 1509
QY 2263 SARPOFOTOPSPAPPEVOTOPEVOTOTVSSHVPSAQPTHAOSKPOVAOSOPQSNV 2322
Db 1510 EOTPNASQOEINDAKQEVTELN-QAKTINDOSTDEYDNAVKEGKAKINA-VKTFSEY 1567
QY 2323 OGOSPVVOSPSQTRI-----RPSTPSQLSPQOSQV-----OTTTSQPIPIQ 2365
Db 1568 KKDALKAKIEAAVNAKYTEADNSWASTSEIAEKOKLAEIKOTADQNVNATSKDDIEVQ 1627
QY 2366 PHTSL-QIPSOQOPQOPQVOSTQTLSSGQTLNOVSVSSPSRQLOIQOPQOVIAPVQ 2424
Db 1628 IHNDLINDVYITPTGKKESATTDIYAVADOKKNNISADT-NATODEKO-----AIKO 1680
QY 2425 LOQOQVO-VLSQIOSQV---VAQIQAGQSGVPQOKILOPI---QIOOSSAVOTHO-I 2473
Db 1681 VQONVOTALESTINGVNDGDVDALTOGKAIDAIQVDATVVKANQVIDAKAEETKEST 1740
QY 2474 QNVVTVQAASVOEQLOQVQOLRDOQ-----QKKQOQTEINVTSPSKLLIKV 2520
Db 1741 DOSDQLTAEKTEALAMIKQITDOAKGITDATTTAEVEKAKAGLEAFDN-----IQI 1794
QY 2521 EIIQOVVAKHNAVIEHLKOKKSMTPAREENORMIVCNOVMKYIILDKID----- 2570
Db 1795 DSTEKO-----KAIEBELT-----ALDOIEAGVNVADADAT 1824
QY 2571 KEKKA-----AKKRKEESVEOKRSKONAT-KISALLFKHKEOLRAE---ILKRA 2618
Db 1825 TEKEKAFNTALEDISKATEDISDQTTNAELIAVKSAL---EOLKAORINPVVKKNA 1879
QY 2619 L-----LDKDLOIEVOEELKRDLLK-----KEKDLMLAQTATAVAPCPPV 2660
Db 1880 LEAIREVNVK--QIEIIRKADADASAKELARTDILGRFPDFADKLDKQTNTEVALQNV 1937
QY 2661 TPVLPAPAPPPSPPPPGVQHTG 2684
Db 1938 T--IPAIEAIVPONDPNANDTNSG 1959

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comogen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:21:41 ; Search time 21.9763 Seconds
(without alignments)
3892.037 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971
Sequence: 1 MVSEEEEDGDAEETDSE.....KLKGFKASHNNKLOSTAS 2907

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	982.5	6.6	238	4	US-09-257-179-80
2	491	3.3	2137	4	US-09-134-001C-4463
3	441.5	2.9	3686	4	US-09-134-001C-5080
4	391	2.6	10182	4	US-09-134-001C-3159
5	367	2.5	3969	4	US-08-061-376-5
6	362.5	2.4	2842	1	US-07-741-940-7
7	362.5	2.4	2842	1	US-08-289-548A-7
8	362.5	2.4	2842	1	US-08-452-654-7
9	362.5	2.4	2842	4	US-08-449-731-7
10	362.5	2.4	2843	1	US-08-452-655B-2
11	362.5	2.4	2843	1	US-08-452-655B-7
12	362.5	2.4	2843	1	US-08-450-582-2
13	362.5	2.4	2843	2	US-08-450-582-7
14	362.5	2.4	2843	2	US-08-821-355A-7
15	362.5	2.4	2843	2	US-09-003-687A-7
16	362.5	2.4	2843	1	US-07-741-940-2
17	361	2.4	2843	1	US-08-289-548A-2
18	361	2.4	2843	1	US-08-452-654-2
19	361	2.4	2843	2	US-08-370-235A-2
20	361	2.4	2843	4	US-08-449-731-2
21	361	2.4	3248	4	US-08-353-700-1
22	354.5	2.4	3248	5	PCT-US95-16216-1
23	354.5	2.4	2035	1	US-08-046-585-5
24	352.5	2.4	2035	5	PCT-US93-11721-5
25	352.5	2.4	2035	5	PCT-US93-11721-5
26	352.5	2.4	2035	5	PCT-US93-11721-5
27	341.5	2.3	2482	1	US-08-328-254-6

28	328	2.2	1852	1	US-08-425-061-24	Sequence 24, App1
29	328	2.2	1852	2	US-08-825-886-24	Sequence 24, App1
30	328	2.2	1863	1	US-08-425-061-16	Sequence 16, App1
31	328	2.2	1863	1	US-08-480-784-2	Sequence 2, App1
32	328	2.2	1863	1	US-08-483-553-2	Sequence 2, App1
33	328	2.2	1863	1	US-08-487-002-2	Sequence 2, App1
34	328	2.2	1863	1	US-08-483-554B-2	Sequence 2, App1
35	328	2.2	1863	1	US-08-798-691-4	Sequence 4, App1
36	328	2.2	1863	1	US-08-488-011B-2	Sequence 2, App1
37	328	2.2	1863	2	US-08-825-886-16	Sequence 16, App1
38	328	2.2	1863	3	US-08-825-487A-4	Sequence 4, App1
39	328	2.2	1863	4	US-09-074-476-6	Sequence 6, App1
40	328	2.2	1863	4	US-08-850-727-2	Sequence 2, App1
41	328	2.2	1863	5	PCT-US95-10203-2	Sequence 2, App1
42	328	2.2	1863	5	PCT-US95-10203-2	Sequence 2, App1
43	328	2.2	1863	5	PCT-US95-10203-2	Sequence 2, App1
44	326	2.2	1863	2	US-08-603-753D-2	Sequence 2, App1
45	326	2.2	1863	4	US-09-099-753-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-09-257-179-80
; Sequence 80, Application US/09257179
; Patent No. 6410709

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 29 Human Secreted Proteins

; FILE REFERENCE: P2015P1

; CURRENT APPLICATION NUMBER: US/09/257,179

; CURRENT FILING DATE: 1999-02-25

; EARLIER APPLICATION NUMBER: PCT/US98/17709

; EARLIER FILING DATE: 1998-08-27

; EARLIER APPLICATION NUMBER: 60/056,270

; EARLIER FILING DATE: 1997-08-29

; EARLIER APPLICATION NUMBER: 60/056,271

; EARLIER FILING DATE: 1997-08-29

; EARLIER APPLICATION NUMBER: 60/056,247

; EARLIER FILING DATE: 1997-08-29

; EARLIER APPLICATION NUMBER: 60/056,073

; EARLIER FILING DATE: 1997-08-29

; NUMBER OF SEQ ID NOS: 128

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 80

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-257-179-80

Query Match

Best Local Similarity 78.8%; Pred. No. 1.3e-51;

Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;

-----DT 2727

QY 2689 PTPVSAQKKRRERK-----DSSSKSKKKKMTSTSKTKK-----

DB 3 PTR-----KTPYDSKRYICDCTNMYHGEVGIPEKAKKMDVYICNCKRAQSSSE 58

QY 2728 KLYCICKTPYDSKRYICGCRCONMYHGRVGIQSAELIDVYCPQCSSTEDATVLT 2787

DB 59 ELVYCICRTPYDSKRYICGCRCONMYHGRVGIQSAELIDVYCPQCSSTEDATVLT 118

QY 2788 PLTEKDYEGKRYLRISQAHKMAPLEVPDNDAPYVGVIRKPMPLATMEERVORRY 2847

DB 119 PLTEKDYEGKRYLRISQAHKMAPLEVPDNDAPYVGVIRKPMPLATMEERVORRY 178

QY 2848 EKLTPEVADMTKTFDNCRRYNNPSDPEYQCAEYLESEFFVOKLGFKASHNNKLOSTAS 2907

DB 179 EKLTPEVADMTKTFDNCRRYNNPSDPEYQCAEYLESEFFVOKLGFKASHNNKLOSTAS 238

RESULT 2

1595 VSTSVSDSTASTSEASSTSTSVSDNSASTSLSESTSTSLDSTSMST-----1643
QY 1971 FHQFTWVWQOGSNGGVVOQOKVLGIIIPSGTSGOQFTSPQPRAT-----VTIR 2023
Db 1644 -SBSASTSTSESDSDASTSLDSTSTSVSESTSTSTSTSVSASNSNSTSLDSTSRSTSL 1702
QY 2024 PNTSGS---GCTTNSQVITGPQIRPGMTVIRPLQOSTLGKAIIRTPVWVGAPQOVM 2080
Db 1703 DSTSTSTSEGSTSTSE-----SDSDASTSLSEST-----S 1734
QY 2081 TQIIRQPVSTANSAPNTVS---STPGQKSLTATSTST-----NIOSSAQPRPQOQV 2132
Db 1735 TSTSDSTSTSDASTSMKSVSDNSASTSLDSTSTSVSDSTASTSEASTSTRESES 1794
QY 2133 KLTMAOLTOLOGHGNGOGLVVIQOGGOTTGOLQILPGCVTLPGGQLOMAAMPNGT 2192
Db 1795 TASTSTSESTSVSDSTSTSTSDASTSTSESDNSESTSLSESTSVSDST-----1850
QY 2193 VQRFLEPLATVATTAATTTTAVSTTAAGTGEORQSKLSPQOVHODKTLPPAOSSSGVP 2252
Db 1851 -----ASTSASASTSTSVSDNSAST-----SLSGSTSTSVSD 1883
QY 2253 AKAQPTAQSARPPQOTQOPQAPQEVQTPQVQOTVYSHVPSAOTTHAQSSKPY 2312
Db 1884 STSTSTASASTSTSESDSDASTSLSGSTSTSTSDSTSTSDASTSTSEASTSTSV 1943
QY 2313 AAOQOPNOVQOGSPYRVOQSPOT---RIRPSTPQSLSGQOQOYQTTTQPIQPIQT 2368
Db 1944 SEDSDSTSVSESSSTSVSDSTSTSTSEASTSTSESESTSESTSVSESSSTSTSDSST 2003
QY 2369 SLOIPQOGPOPOVOOSTQOTLSGOTLMOQVSSPSR 2407
Db 2004 STSMSTSEFTSQSPINSEQFT---GDSLEDTITVQSK 2040

RESULT 3
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stramm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.9%: Score 441.5; DB 4; Length 3696;
Best Local Similarity 17.1%; Pred. No. 2.1e-17;
Matches 471; Conservative 470; Mismatch 1066; Indels 749; Gaps 106;

QY 311 QAKKPYIRHEPICYD-RSRKKYWFNLNR--LITEEDTENENEKKIWWYSTKVOALAEIDLC 367
Db 569 QVAKDIL---PSNYTLASVYKYNKIKERAKQTVLDEETNNP-----606
QY 368 LKQDYEAELCKLLEKREIHRHMDITDITNKARGSKNSPLAANEELLESIRAKKD 427
Db 607 FNGRYSQTOIDDLIHEIQTLLIRVSAEREINKA-----QEMTDAY-----648
QY 428 IDWVKSPEEFEXKKNFENDSKDAEKNREFEEDQSLEKSDDKTPDDPQCGK-----480
Db 649 -----YSTELTTEKDTLVQILENHKNEISNNIDELTIDGVERKVEAGLHTL 697

QY 481 -SEEPTEV-----GDKNSYSANLGDNTTATSE-----ETSPSEGRSPVCL 522
Db 698 ESDTPHPVTPKPNAROVYNNRNDQOKTLIRNNHEATTEEQNEAIRQVFAHSDAIKIG-E 756
QY 523 SEFPDSSNMAEKVYASLTPDQVEEPKTKCESSNTASTTTSIQPNLENSMSSE-----576
Db 757 AETDTTYVNEARDNGTKLIADV---PMPYTKAARAAVNTSANSKIKDINNNTQATLDER 813
QY 577 -----LNSOSE-----SAKAD-----PENG-----ERESHTPVSIQBEIV-----609
Db 814 NDAIALVNRKSDALQININAGNDVTEQNNGTNTIQVPLTPKVRMAIATINAKD 873
QY 610 -----GDFTEKSTGELSESPGAGSGSTRIITRLNPPSKLSQLKSQVAAA 659
Db 874 EOKRLIOANNATTTEKADAEKRYNEAV---ITANQITNATTNRVDQAQTGGSGIISA 930
QY 660 AHEANKLFEKGGK-----EVLYVNSQGEISLSTKKEYIMKGINNNFKLGQEGKYRYHN 714
Db 931 ISPATKIKEDARAAVEAKAIAQNOQINSNMATTEE---KEDALNDVEARKQAATINQ 987
QY 715 QYSTNSFALNK-----HOHREDHDKRRLAHKFCITPAGFEKRWNSVHSGKVLITSTL 767
Db 988 AOSTQOYSEAKNNGINTINDQOPNAVKKNT--KTILEQGNKR-KSAIQOTPDATTEEK 1044
QY 768 RLTTTOLENNIPSSFPHNASHRAMWIKAVQMSKREFALALILECAVKPYMLPIW 827
Db 1045 QEAVSAVQAQVNTGITHINQANSND-----VDQELSNAEQIITHTNVNVQKKPQA 1095
QY 828 REFLGHRRLMRHSIEHEKEKVKKKK-----KQEEELMQOATWVKYFPYKHQV 879
Db 1096 ROALIAKTNEKQASINDNGTILEKQAKIQSINDAANIADDEQITQAA-----1143
QY 880 WKQGEERYRVYGGWMSIKTTHYRFPKLPQNTVNYRKSLSEGTNNMNDENDESCKR 939
Db 1144 ---SNQVNDALNIGISINIKIOT-NFTKKQOARDQVQNFQKEKELNTPHATQDEKQ 1199
QY 940 KC---SRSPKIKIEPDESEKDEYKSDAKGADQENEMDITEKKQDVYKELSDSDK 996
Db 1200 DALTRLQAKETALNDINOQOTNOMVDTALTSQIONQIONQVNVKQKQAKTTI-NDIYQ 1258
QY 997 PCKEEPHEVDDDKTBESHV--NQOESSQVNVN-----YSG-----FHL 1034
Db 1259 QHKQSIQONNDATTEKEVEVANNVMAQOVNSKIDNATTNNOIDIGISGRQSIATP 1318
QY 1035 RTSYK-----KTKSSKLDGLE-----RRIOFTLEKORLEKLEGG 1074
Db 1319 DSTIKRRAKNDIDIKADKKIKIQRINDATDEIQAENKRIEPAKIPAKINQNRSTQ 1378
QY 1075 IKGIGKTSTNSSKRLSESPVY-TRAKEGQSDSMRQ---EQSFMANNQPEDLIQGCQ 1129
Db 1379 VNEAKTNGIKIKIENITPATTVKSBARQAVONKANQEIINHQINPDATNEKQKAINVSA 1438
QY 1130 SDSSVLRMSDPSHTTNKLYKDKRYLD--VSIRSPETCKQNSIENDIEKYS-----1181
Db 1439 ELARVOAQIINAEHTQGV---KTIKDQATISLRINAQVVEKSSARARAIQKATQOTFT 1495
QY 1182 -----DLASRQDEPTKSK-----TRKGNDFPIDSKSLASADDIGTLI-----1217
Db 1496 NNNDNATDEKEVANNVITKQKSLDINISLSMND--VENAKVAGINIAVLPAITAV 1553
QY 1218 -CRNKPLIOESDPTTVYSSSKALHSSVPKSTNDROT-PLSRAMPFEGKLGDSF--SN 1273
Db 1554 KSKAKKIDQKLAQOI-----NQIQHTQATTEKEKAQIAQLANQKSSEARTAIQNEHSNN 1608
QY 1274 STLENSSDIYS-----IQDSSSEDMIVQNSNESTSEQF-----TRQDYVLEPLK 1320
Db 1609 GVAQAKSNGIHETELVMPDARKKSDAKOSIDNKRYNEQSNTINTTPATDEKQKALDKIK 1668
QY 1321 CELVSG-----ESTGNCEDRLPYKGTGAN-CRKPQOQKLEER-----1357
Db 1669 IAKDAGINKVDAQOTNOQVSDAKTEALIDITITNIOANVAKPSARVELDSKFEDLKROI 1728


```

QY 1358 -----PVKCSDOIKNKTDDKNNNESEKKG-ORTSTFOI----- 1394
Db 1729 TPNAITEEKODAIORLNGKDEYKNNLINODRDRNEVEQHHNIGQLELTIHANPTKSDA 1788
QY 1395 -----NGKONKRYIYKGCLEKISESVVSGNVEPKVNN----- 1429
Db 1789 LOELOTFISQOTELINNNKDATNEKDEAKRLLEISKKTITTNINQAJTNQOVNADKNG 1848
QY 1430 -----INKIIPENDIK-----SLTYKESAIREFING--DVIMED-----FERNSESEKSH 1473
Db 1849 MNEIATITIPATTTIKDAKTIDKKAEOQVITINGNNNDATDEKAEARKLVEAKKIEKSN 1908
QY 1474 LLSSS---DAEGNVRDLETLPTSKESDSTQTTTPSASCEPSNVNOVEDMEIETSEVK 1530
Db 1909 ITNSDTEREVNKAKTNGLEKINNIOPTQTKT-----NAKQJINNAQE--QLIO 1956
QY 1531 VTSSPTISEESNLSDNFIDENGLPIKKNENNGESKRKVTVEVTMTSTVATESKTVI 1590
Db 1957 INMTPATEEKEKEATIRVNA-GL-AQAIQININNAHSTQENESKTSIATIKSVQPNVI 2014
QY 1591 K-----VEKGDQOTVVSSTEN-----QAKSVT--TTTTVTKLSTPSSGSDI 1633
Db 2015 KKPRAINSLOEANNQKTLIGNOGNATDEKKAQOLVTKLNEQIOKIHESQODNOVD- 2073
QY 1634 ISVKEOSKIVVTTTVDTS-----LTTTGGTLVTSMTVSKEYSTEDKVLKMKESRPK 1684
Db 2074 -NKAQAIFTAIKLINNAHNRQDAINTLTLAESKSDIHANDATTEEK----- 2122
QY 1685 KTSNGTALPSYKRFVSTKKSIFVLPNDLKLARKGGIREVPYFNNAKPALDIWYPY 1744
Db 2123 -----NTAIOSTI-----DDTLQAR-----NNINGANTNA----- 2147
QY 1745 SPRPTFGITWRYLQTVKSLAGVSLMLRLMASLRMDMAKYPGGSGSTRTESEIT 1804
Db 2148 -----LVDENLBDGKOKLQRIYLSVSTQTKQAADIAQAOQSRSTIDQONA 2194
QY 1805 TT-----EITR-----RDVGPYGI--REFEYCI 1825
Db 2195 TTEKQKALERLNOETNGVNDRIQALALANONVTDEKNNILETIRNVEPIYIVKPKANEII 2254
QY 1826 RKIICPIGVEETPREPTPOKGRSSALRPKREPTEKQGVPIIETWVAEELMEIR 1885
Db 2255 RKRAEOTTLINONQATLEKQALGKLEVKNEALNO-----VSGAHNSNDVKIAENN 2309
QY 1886 AFAERVEKEKAQAVEQOAKRLBQOKPVIATSTSTSTSTSPAQVMAVIS--G 1943
Db 2310 GIAKISEVHEBTIIRKNAKQIEBQDAQSQIDITINANKSTINEKSAIDRVNAKIDAIN 2369
QY 1944 SVTTGTAMVLTITVGSFAIYTFQONKNFHOTFATWAKOGSNGSVVOYQKVLGIIPSS- 2002
Db 2370 NITNAT-----TQQLVNDV-----KNSGNTSISO-----IIPSTA 2399
QY 2003 --TGTSQOTFSOPRATVATIRPNTSGSGGTTNSQVITGCPQIRPQMTVITRPLOO-- 2057
Db 2400 VKINATAALALASEAKKNKALIIDOTPNATAEKEKANNV-----DRLOEAD 2445
QY 2058 STLGAALIRTPV--MVGPGAQO--VMTQIIRQOPVSTAVSABNTVSTPQOKSLTSATS 2113
Db 2446 ANIKAKHTDEVNNIKKQAOVINAVQEVIKKQVKNQJN-----QFIDNQKIIENTP 2500
QY 2114 TSNIGSASPPRPQOGVULMAQLTQLTQGGGNOGLTVIYIOGSGQTGQIQLIPQV 2173
Db 2501 DALLEEKA--EANKLLOWLVLSTSDIELANVDHNEVDQALD-----KARPKIE 2546
QY 2174 TVLPGPOQMOAMPNGVQREFLFPPLATTASTTTTSTVSTAAGGEOROSKISPP 2233
Db 2547 EIVP---QVSKKDVLAIOEAF-----NSQOTIEONQENATEEKEALNKI 2591
QY 2234 MQ-VHODK--TLPPAOS--SVGPAKQOFQIOPASAPQOPQOPASQAPQEVQOTQPEVOTQ 2290
Db 2592 NQNLNQAKVINIDQASQKVDASAKTISIODIEDIQPHQPKATGRHNLNKAN--QQQS 2648
QY 2291 TVSSHVPSEAOPTHAOSKPO-----VAAOSQPOSN-----VOGQSPVRVQSPSQTR 2337

```

```

Db 2649 TIATHPSTIEBROEASAKIQEVLKAKIAKIDKQGTNDVDEKTVVNGIAEINLEIPA-TT 2707
QY 2338 IRPSTPSQLSPGOQSOVOTTTSDPIPIOPHTSLQIOPSQOPQOPQOVSOTQLSSQTL 2397
Db 2708 VKDKAKADVNAEKEOK-----NLQINSNDATTEEEKLAVSDNLNHVET 2752
QY 2398 NOVSVSSPSRPOLOIOP-----QPOVINVPLOQOVVYLSQOSOVVNAIQOQS 2448
Db 2753 NQALIEDAPDTONVAVENKNGKIGITRIQIPLVKKPTAK-----SKIESAV----- 2797
QY 2449 GVPQOLKLOPLIQOQSSAVQTHQIQNVVTVQAASVOEQLQRYOUL-----RDQOQKKQ 2503
Db 2798 -----EKKTEINQOTQATHE--VEGJNQJNQJHEKAKNDVNSQTN 2839
QY 2504 QOIEINVTPSKLLIKYEITIQOVYKHNAY-----IEHLKOKKMTAERE-- 2550
Db 2840 QOYE--NABONSIDQINNRPDPFSKRRNAVAEIVKAQOKIDEIEQEFSAQOEKNAL 2896
QY 2551 -----ENORMIYCNQVMKY-----ILDKIKKEKQA 2577
Db 2897 QHLDQYKEIINSTINCANTDNEVDNAKTSGLNNITTEYRPEYNNKKNAIILKYVSDIQEA 2956
QY 2578 KRRKRESVBQKSKQNAKRLSALLFKHKBQLEAEILKRRALLDKDLQIEQBELK 2633
Db 2957 IINGYPATEDELOEAN-SKLNKILLDKAKQIGIAHTNNVE--DDIYNEVSQKMK 3008

```

```

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

```

```

Query Match          2.6%; Score 391; DB 4; Length 10182;
Best Local Similarity 16.9%; Pred. No. 9,9e-14;
Matches 500; Conservative 450; Mismatches 114; Indels 894; Gaps 116;

```

```

QY 287 EDMQCEVCYAHKVPVGTDCVAELIQNKPYIRIEPIGYDRS-----RRKYFLNRLI 340
Db 2666 KEOLQOSIQPAASTGDGTQDSTRNKKRQAQAQAIQAHANSVYINNGATSQQINDAKNTV 2725
QY 341 EEDTENENNEKKIYYSKYVQYLAELIDCLDKDY-----EALCLILEMREIHRHDITE 396
Db 2726 EQAQRDVEAKSNLRADKSOLOSAYPTLNRDVLTLNDKKPASVRVYDAISNIKEELDTAK 2785
QY 397 -DLTNKARGSNKSF-----LAAANEELISIRAKKGDIDNVKS 433
Db 2786 ADASSITLRNTPVEVEQYRDALNKINTYQPKVNOAIALLQPKENNSSELYQAKKRLQDAVND 2845
QY 434 PEETEKDKNETENDSKDAEKNREE--FEDQSLEKSDDKTPDDDPQGSSEEPTEY---- 487
Db 2846 IPOTQGTQOTINNVDKQREARALTSAGRVINDGDATQOETTSBKSKVEQAMQALITNA 2905
QY 488 -----GDKG-----NSVSANLGDNTNATSEETSSEBSRSVYGCISFPPSSNMA 532
Db 2906 KSNLRADKNELOTAYNKLIEVNSTN-GKKPASIRQYETAKARLQONIN-----DAKNEA 2958

```



```

Db 4909 -----LIPN-----LTFWQAKAKEELVNSQTRREV 4933
QY 2233 QMVOHQDKTLPPAOSSSSYGPAAK---AOPQTAQPSARPPQIOPSPAPQEVQTOPEVQTO 2289
Db 4934 QEOLNQAASL---DSSGSLKSLVAKQPIVQKTSVYINEDQPPSOAVNDSIT---MGQ 4985
QY 2290 TTVSASHVEAOPTHAOSKSKQVAASQPSQSNVQSGSPFVQSPQTRIRSTPSOLSFG 2349
Db 4986 TIINKTADPVLDKTLVNDNAISINISTKENALHGEQKLTAKTEALNAL---NTLADLNTF 5041
QY 2350 QOSOVQTTTQOPPIQPIQTSIQIPSOQPOSOP-QVOSSTQTLSSGQTLNVOVSSPSRP 2408
Db 5042 QKEAIKTAIN---TAHTRFDVTAE---QSKRANOINSAMHTLQNSIDNE-SVTNEN- 5091
QY 2409 QLOIQOQPOQVIAPQIQOQVQVLSQISQVVAQIQAOQSVPOQIKQPIQIQOQSAV 2468
Db 5092 -----YNAEPEQHAFTALNNAKETVNE-----QOATLDANSINOKQAOI 5133
QY 2469 QTHOIONVTVQAAVQEOLOQVQOLRQOQKQOQOIEINVTPSKLL---IKVELIQ 2524
Db 5134 LT-----TKNALDGEQLRAKENADQELINTLOQLDQNRSEKGLINSQTRTEVAS 5186
QY 2525 KOVVMKH-NAVIEHLKQ---KKSMTPAER---EENQMTVCNQVMKYIIDKTDKEEQ 2575
Db 5187 QLAQAKELNKVMEQNLNLINGKNOMINSKFINEDANQOQVSNASAEALK-NKSQNP 5245
QY 2576 AAKRRKRESEVEQKRSKONATKLSALLFKHKEQLRAELIKRALLDMDLOIE---VOEEL 2632
Db 5246 ELDKVTIEQALINNINSALINNNGEAKLTKAKEDAVASINNLSGITNEQTKENOAVNQAQ 5305
QY 2633 KRDLIKKEKDLMOJQA 2650
Db 5306 TRDQVANLRDAEALDQS 5323

RESULT 5
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djahail, Malek
; APPLICANT: Sellerl, Licia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; TITLE OF INVENTION: TRANSLLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEO ID NO: 5
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

Query Match          2.5%: Score 367; DB 4: Length 3969;
Best Local Similarity 17.6%; Pred. No. 7e-13;
Matches 620; Conservative 405; Mismatches 1211; Indels 1278; Gaps 153;

QY 5 EEEEDDDAEETQDSEDEDEMEDEDDSDYPEMEDEDDDDASYCTESS-----54
Db 855 EELSKRDADQSVKDKSRERDREREKENK--RESRREKRRKGSEIQSSALYVGRVSK 912
QY 55 -----FRSHSTYSSTPGRKKPRVH-----73
Db 913 EKVVGEDVATSSSAKAKTSSSHDSGDTITSVLGDTTAVKTLIKKRGONLEKTN 972
QY 74 ---RPRPILEEKDIPLEPFKSSSEDLAVPNEHIMNVIATIEVLRNFGTVLRSPFRED 130
Db 973 LDLGPTAPSLKEKTLTSLTPSSS-----TYKHSITSSIGSMU-----1009
QY 131 FCAALVSQEOCTTMAEMHYVL-----KAVLEEDTSNTTF- 166
Db 1010 -----AQADKLPTWTDKRVASILKKAKAQOLKIEKSKLKOTDQPKAQGOQSDSSETSVR 1063
QY 167 GP-----ADKDSVNSTLYFLDGMTPPE---VLRVYCESDKREYHVL 206
Db 1064 GPRIKHVCRRRAVALGRRKRAVFPDMDPT---LSALPWEREKILSSMGNDSS--SIAG 1117
QY 207 YQEAEDYPYGVENKIKVLOFLVDQFLTTNARELSSEGVIOYDHCRCYHKLGLDLC 266
Db 1118 SEDAE--PLAPPIKPIK-----PYTRKAKAQEPVKKGRSRRGQ-----C 1157
QY 267 ETCSAVYHLCVKPRLPEVPEDEMOCVCAHKKVGVTCVAELIQKKPYIRHEPIGYDR 326
Db 1158 PGC-----QVPED---CGVC-----TNCU-----DKP-----1176
QY 327 SRRKWFNLNRLIIEEDTENENKKIYVSTKVQJALDLCLODKVYEAFLKILEMRE 386
Db 1177 ---KFGGRN-----IKKOCKMKKCONLQWMP 1201
QY 387 EIHHRMDITEDLTNKARGSKSFLAANBEILSIRAKGDIDNVKSPEETEKDKMETEN 446
Db 1202 KAY-----LQKQAKAVKK-----KEKSKTSKSKKSSSVKKNVDS 1239
QY 447 DSKDAENKREFEEOLEKSDDKTPDDDPQOKSEPETEVGDKGNSVSNLGDNTNAT 506
Db 1240 SQKTPSARBD---PAKKSSSEPPPRKPVEEKSEB-----GN-VSAP-GPESKOAT 1286
QY 507 --SEETSPSEGRSPVGLSETPDSSNMAEKKVASELPODVPBEPNKTCESSTNTSATTT 564
Db 1287 TPASRKSSKVQOPALYIPQPPPTGPPRK-----EYKTTTSPERKK-----1329
QY 565 QPNLENSNSSSELNSSQSESAAADP-----ENGRESHTPVSIQELIVGQFTSEKST 618
Db 1330 QP-----PPESGPEOSKOKVAPRPSIPVKOKPKPEKEKPPVKNQEN-----1372
QY 619 GELSEPSGAKGASGRITRLRNPDPSKLSQKSOVAAAHAANKLFPREGKEVLVNS 678
Db 1373 -----AGTLNLTSLTSGNS--KOKIPADGVHRIYVDFEEDCAENWVE 1415
QY 679 QGEISRLS---TKKEVIMGNINNYFKLQGEQKYRVYHNYSTNSFALMKHQHRDDHK 734
Db 1416 MGLGILITVPIITPRVVC-----FLASSGHEVEVYQVCCPP-----1454
QY 735 RRHLAKHFCFLTPAGEFKWNSVHGSVLTISTLRLITLOENNIPSPFFHNNASHRANN 794
Db 1455 -----HKEFCE-----ENRPLEDDQLENNCCRC-- 1478
QY 795 IKAQVOMSKREFALALALIEC-----AKPVVYMLPIHREFLIGHTRLHR 838

```


Db 1479 -KCHVCG--ROHOATKOLLECNKNSYHPECLGPNYPKPTKKKKWMI-----CTKCVR 1531
QY 839 MWSIEREEKYKKKKKEKOEETMOOATWVKYTFPVKHQWVK--OKGE-----885
Db 1532 CKSCSGSTTPGK-----GMDAQW-SHDFSLCHCAKLFKAGNCPCLDKCYDD 1577
QY 886 ---EYVTVGGMW-SIS-----KTHYVRFVKKLPGN---ITVYNRK-----920
Db 1578 DDEYSKMMOCGKDRWVHASKCENLSDEMYELLSNLEPSVAAYTCNCTERHPAEMRLALEK 1637
QY 921 ----SLEGTKNMDEMDEMSDKRCSRPRKIKIEPDS-----KDEYKGS-----AA 965
Db 1638 ELDISLKQVLTALNSTRSHLRQAAKPPPLNPETESISTRSRSPEDDPVLEVS 1697
QY 966 KGAQONEMDISKITEKKDD-----VKELSDSDRCKEEMPEVDDMK 1010
Db 1698 KODQOQPLDEGVKRRMDGNNTSVLEFSDDYKIIQAAINSDGGP-----EIKK 1748
QY 1011 TESHVACQSSQVYVNVSEGFHLRTS-YKKTSSKSLDGLERIKQFTL-----1060
Db 1749 ANSMVKSFTROME--RVPPWFSVKRSFWEPRKVSNSGMLPNAVLPSLDHNYAQOWE 1806
QY 1061 -BEKORLEKILEGGI-----KGIGKTSNSKNSLSEPVITKAKEGQSDSMROQSP 1113
Db 1807 REENSHTEOPPLMKKTIIPAKPKGPEPDSPTPLHPPTPIL-----STDRSKEDSF 1858
QY 1114 NAN-----NDOPEDLIQ-----CSQSDSSVLRMSD-----1139
Db 1859 ELNPPGIEDNRQALCLFYGDSDSANDAGRLYIGQNMETHVNCALMSAEVEFDDGSLK 1918
QY 1140 -----PSHTTNKLYPRDVLDDVYSINSPEKCKRONS 1171
Db 1919 NVMMAVIRGKOLNCEPCQPGATVGCCLTSCSYNHFMCSSFRANCVFLDKKVKYCORHRD 1978
QY 1172 I-----ENDIE-----EKVSDLAS--RGOEPTSKTKGNFIDDD-----SKLAS 1209
Db 1979 LKGEVVPENGFEVFRVYVDEFGISLRKFLNGLEPENIHMMIGMTIDCGLDNLDS 2038
QY 1210 ADD-----IG-----TLICK--NRKPLIQ-----EESDITVSSS 1236
Db 2039 CEDKLPPIGQCSRYVWSTTDARKRCVYCKIYECRPRVPEPDINSTVEHDEKRTIAHSP 2098
QY 1237 KSMHSSVKSSTN-----DR-----DATPLSRAMDF--BEKLG 1268
Db 2099 TSTETSSKESKQNTAEIISPPSPDRPHSOTSQSCYHVIKVPRIPTESYSPTORSPGC 2158
QY 1269 D--SESNSTLENSSDTVS:IODSSEDMIYONSNESTSEQFRTREQOVLEP--LKCELY 1324
Db 2159 RPLPSAGSPTPTTHETIYVGP-----LSSGLRSTIG--SRHSTSLSPORSKLRTM 2209
QY 1325 SGESTGNCEDRLPVKGTENGKKRPSQOK--KLERPVNCSDQIKLNTTDDKKN-----1377
Db 2210 SPARTGNFYSRNWVSVSSTGTATDESSAKYVDHGLPLNSTSLGQNTSTSNLORTY 2269
QY 1378 ---ERNESKKGQRTSF-QINGKDKPK-IYKGECLKEISRYVSGNVEPKVNN---1429
Db 2270 VYVGNKNSHLDSSSEMOKSSASDLVSKSSLSKGEKTKVLS-----SKSSESSAHNVAY 2324
QY 1430 --INKLIPE-NDIKSLTVKESAIIRPINDGVIMEDFENRNS--SEKSHLSSSDAAGNR 1485
Db 2325 PGIPLKAPVYHNTSRELNVSKISGFAEPSSV--SFSSKEALSFPHLHLRGQRNDQDHT 2382
QY 1486 DSLETLPTSKESDS--TOTTPPSACSPESNSVNO-----VEDMEIETSE 1527
Db 2383 DSTQANSNSPDDTEVKTLKLSGMSNRSSIINEMHSSSRDRROKKCKCKEJFKKHSS 2442
QY 1528 VAKVYSSPTSEESNLSDPTDE-----NGLPINKNE 1560
Db 2443 KSFLEPGVYTGEEGNLKEPEFDEVLTPRYMGORPCNNVSSDKIGDKGLSMGCVKAPPM 2502
QY 1561 NNGESK-----RKTVIEVNTMTSTVATESKTVIKVKGQKQYVSSSTENAKSTVIT 1614
Db 2503 QVGBGAKELQAPRKRTVYVTLTPLKMNESOSKNALK-----ESSPASPLOIESTS 2553

QY 1615 TTTVTYKLTSPSTGGSVDDIISVEKQSKTVYTTVTVDLSLTGTLVTSMTVSKEYSTRDK 1674
Db 2554 PTEPISASENPBGD-----PVAQPSPN--NTSCQOSQSNNTQNL-----PVQDR 2595
QY 1675 VKLMFSRKPKKTSSTGALPSYKRFYTKSTKKSIFVLPRNDLKL--ARKGIREVYVF--1730
Db 2596 -NLMPLDPKPOEDGSEFKRRYPRRSARASNMFE-----GLPPLGVRSYGEEDIPFYSS 2649
QY 1731 -----NYNAKPALDIMPYSPSPPTFGITWRVRLQTVKSLAGVSLM 1771
Db 2650 STGCKKRGKSABEQVGCADDLSTSEDDDLTYYNFT-----TVISSGGE-----2694
QY 1772 RLMLASLRMDMAKVP-----GGGSTRTETSETTEITTLIRKRDVGYGRIFFEYCIRK 1827
Db 2695 RLASHNLFREECODLPKISQDLGDVGDTE-SDTSVTAT-----2732
QY 1828 IICPIGVPEPKETPPRQKGLRSSA--LRPKRPEPKQTPVLIETWVAEELMEIR 1885
Db 2733 -----TRKSSQIPKRGKNGTENLKDIPEDAGEKEHNTKSSVGHKNPKMNDCH 2783
QY 1886 APFAERVEKKAQAVEQAKKRLQOKPVIATSTSPSTSTISIPAKVWVAPISGSV 1945
Db 2784 SVSR-----VKTOGODSLEAQLSSL-----ESSRHHNTSTPSDKNLD-----2821
QY 1946 TTGKKNVLTTKVGS-----PATVTFQONKNFHQTPATWVWQOSNS-----G 1987
Db 2822 TYNTTELLKSDSDNNNSDDCGNLLPDSIDIMFVLKNTPSMQALGESPPSSSELLNLGEGIG 2881
QY 1988 VVOVOOKVIGIIPSSSTGTSQOFTSFQPRATVYTPRNTSGSGGTTSSNOVITGPDIRG 2047
Db 2882 LDSNRKDMGLF-----EVSQOLPTEPEV--DSSVSSISABEQFELPLEPSPD 2929
QY 2048 MVIYTRPLOGSTLGAIIIRPVWVWPGAPQOVMTQIIRQPVSTAVASANTVSSTPGOKS 2107
Db 2930 LSVLTT-----RSPVY-----PSQNPRL-----AVISDGEKR 2958
QY 2108 LVSATSTSNIOSSASOP-----PRPQ-----OGUV--KLTMALOLTQOGHG 2147
Db 2959 VT--ITEKSVASESPPALLSPGVDPPTREGHMPRHFIQGHMDADHISSPCGSVQGHG 3016
QY 2148 GNOGLTVILOGGQTTGQIQLIPQVTV-----LPGQO-----2181
Db 3017 NMODLT-----RNSSTPBGLO--VPSVPTVIONOKYVNSTDSPGQISNAAVQTPPHL 3070
QY 2182 -----OLMO-----AAMPNGTVORFLETPLATATATSTTTVYVSTTAAGTGHQ 2225
Db 3071 KPAETELIYVNOQMOPLYVLQILPKNVYTKIQLTSSVSSTPSMETINTSVLGPMSG--GLT 3129
QY 2226 ROSKLSPOQVHODKTLPPAOSSSVGPAP-----AQPTAOPSAAPQOPOTP-----2272
Db 3130 LFTGLNPLSLPTSG--SLPFSASKGLLPMHGHHLHFPATQOSPSPNINSPGSLIGV 3187
QY 2273 QSPAQPEV--QTPREVQOTIYSSHVPS-----EAOP-THAOSKSPVAAOSQPOSNAVQO 2325
Db 3188 QEPEDQOLLVESQRTDSTIYVAT--PSSGLKRPISRLQTFKKNKLPASSPSPNI---3242
QY 2326 SPYRQVOS-----PSQORIRSPSTPSQSPG-----QOSQOVQTTTQOPIP 2363
Db 3243 ABSQDVVSNMNTLINFPTSQL--PNHPSLLDLGSLNTHSHRYVNTIIRKKSSTIMTFEPAP 3299
QY 2364 IOPH-----TSLOIPSOQPOQPOVOSSTOQLSSGOTLNOVSVSPSRPOLQIOQP- 2415
Db 3300 LLPQSVGTAATAAGSTISQDPSHLTSSVSGLASSSSVLNVMSQTTTTPTTSSASVYG 3359
QY 2416 -----QPOYIAMPOLQOQVQLSIOQVYVAQIOAQOSQVPOQIKQIDPIQIOQSSA---2467
Db 3360 HVTLPNRLIGTPDIGS-----ISNLLIKASQOSIGIODO-----FVALPSSGMPR 3406
QY 2468 -VQTHQIQNVVYVQAASVQEOQLQVQQLDQOQKKQOQIEINVMTPSKLLKVELIQO 2526
Db 3407 QLGTSQTPSTAATAAASICVL-----DSTQTTGLTAASPS-----3442


```

QY 2527 VAKHNAVIEHLKOKKMTAPAREENOMIYCNQVMKYILDKIDKEEKOAKKRRREESV 2586
DB 3443 -----GEADENHOLOHVNOVL----- 3458
QY 2587 EOKSKONATKLSALLFRKHEQRLAEILIKKRALDKLOLEVOELKRDLIKKEKDLMO 2646
DB 3459 -----ASKTGI-----HSSQ-----RDLDSAGPOVENFTQTVAPRSMG 3493
QY 2647 LAQATAVAAPCPVTPPLPAPRPPPPPPPGVOHGLST--PTLPVASOKKREEEK 2704
DB 3494 LEQKALIS-----SAYQASTSPGSGSPSSGQSRASASVPEPTPKPKTKFQPLD 3547
QY 2705 DSSSKSKKKMIPTSKET-----KKDTKLYCICTPYDESKFYIGCDRCQNMVHGRCVGI 2760
DB 3548 KGNCKKHVSHLRTSSSEAHIPDOETSLNSGTGTPGAEAE----- 3588
QY 2761 LOSAELIDEVPCOCOSTEDAMTVL-----TPLEKDEGLKRVLRSLQAHKAMPF 2813
DB 3589 -QODTASVEOSSQECQAPQAVLPEVQVTONPANEO----- 3626
QY 2814 LEVDPDAPRYGVIVKEPMDLATMEERVORRYEKLFEVADMTKLEFNCRYNPSDSP 2873
DB 3627 -ESAEPTVEEESNFSPPLMLQOQOKK---ESTTEKKPKGLVFE-----ISSDQ 3677
QY 2874 FYQCAEVLESEFYOKLKGFKASRSHNNKLOSTAS 2907
DB 3678 FOICAESIED-----AMKSLTDKQOEAR 3701

RESULT 6
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US-07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids

```

```

; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-07-741-940-7

Query Match 2.48; Score 362.5; DB 1; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8.1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

QY 373 WEALCKILDEMPREIRHMDITEDLTNRKAGSKNSKSLAANEELIESIYAKKGDIDNVK 432
DB 698 WDMGAVMLKLNLIHSHKMT-----AMGS-----AALRLMANRPAKKKDA-NIM 742
QY 433 SPEET-----EKDKN-ETENDSKDAEKNEEFEDQSLKSDDKTPDDPEQKSEPT 485
DB 743 SPSSSLPSLHVKKKALELDAQHLSETFDNIDNLSPKASHRSKQHNKQSLVGDVYFD 802
QY 486 EYCDKGNVSANLGD-----NTT-----NATSEETSPSGRSPVGC--- 521
DB 803 NRHDDNRSNDNFNGMTVLSPYLNTTVLPSSSSRGLSDRSRSEKRLERERIGLGNY 862
QY 522 --LSETPDSSNMAKKVASELPQ--DYPEBN--KTCESSNTSATTTSI-----QPTLEN 570
DB 863 HPATENGTSSKRGLOJSTTAQIAKMEVSAHTSQEDRSSGSTITELICVDERNAIR 922
QY 571 SNSSSELNSSQSESAKAADPENGERSHPVSTQIEIVGDTSEKSTGELSESPGAGK 630
DB 923 RSSAAHTHSMTYNTTKS-----ENSNFTCSMPYA---KLEYKRSSDNLNSVSSSDGYGKR 975
QY 631 ASSSTRILITLRNPDKSLQLS--QQVAAAHAENKL-EKEGEVLVNV----- 677
DB 976 GOMKPSIESTSEDESKFCSTGYQYRADLAHKHSANMDDNGLDTPYSLKYSDEQL 1035
QY 678 SOGEISRLSTKKEYIMKGINNVEFKLQEGKYRYVHNOYSTNSELNKHQREDHDK-RR 736
DB 1036 NSGRQSPSQNERMARPRHIIIEDELKQSEQQR--NQST--YPVYTESTDCKHLKFOR 1090
QY 737 HIAKFCILTPAGEFKKNGSVHGSKVLTITLRLITTOLENNITSSFFHPWASHRANWIK 796
DB 1091 HFGQECVSPYRSRANGS-----ETNRQGS-----NHGIGNQV 1124
QY 797 AYQMSKPREFALALILLECAVRYVMLPIWREFLGHTRLHRMTSIREKEKVKKKKK 856
DB 1125 SOSLCQED-----DYEEDKPTNYSERY 1146
QY 857 QEEETMQOATWVYKTPPVKHQVWKQKGEYRVYTGGMWSMISKTHYRVPRLPGNTNV 916
DB 1147 SEED-----QHDEEERPTNY-----SI 1163
QY 917 NYRKSLEGTNNNDENDES-----DKKCSRSPPKIKIIEPDSEKDEV 959
DB 1164 KYNEE-----KRHVQPIIDYSLKYATDIPSSQKQSFSPKSSQSSKTEHMSSESTNTST 1219
QY 960 KGSAAKGAQONEMDIDKIEKKQDVKELLDSOS-----DKP-CKEERMEV 1005
DB 1220 PSSNAKR--ONLHPSSAOSRSGQPOKAATKYVSSINQETIOTCYVEDYPIJFSRCSL 1276
QY 1006 DDDMKTESHYNCOSSQ-----VDVVNVSEGFHLRT-----SYKKRTSKSL 1047
DB 1277 SLSASADEIGCNGTTOEASANTLQIAELKEKIGTGSADDPVSEVPAVSQHRRTSSRL 1336
QY 1048 DGLIERIKQFTLEEKORLEKIKLEGKIKIGTSTNSSN-----LSESPVI-TRAKES 1101
DB 1337 QG-----SLSSESARHKAVERSSGAKSPSKAGATPRKSPENHYQOETPLMFSTRCTSV 1389
QY 1102 QQSDSMQKQSPNANNQPPEDLIDQCGSQSDSSVLRMSDPHTTNKLYPKRVLDVDSIRS 1161
DB 1390 SLSDFSRSIRSIASSVOSEP-----CSGMVSGIISPSDLDPDSGQWMPSPR----- 1434

```


TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 2.4%; Score 362.5; DB 1; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8.1e-13;
Matches 409; Conservative 327; Mismatches 86; Indels 777; Gaps 98;

QY 373 WEAEIKLLEEMREIHHMDITDLTKKAGSNKSFIAANELLISIRAKGDIDNVK 432
D 698 WDMGAVSMLKMLHSHKMI-----AMGS-----AALRLNMAHRAKYDA-NIM 742
QY 433 SPEET-----EKDKN-ETENDSKDAEKNEEFEDQSLFKSDDKTPDDPEQKSEPT 485
D 743 SPGSLPSLHVRKQKALAEALDAOHLSETFDNIDMLSPKASHRSKQRIKQSLYGYVYEDT 802
QY 486 EVGDKGNSVSNLGD-----NTT-----NATSEPTSPSGRSPVGC--- 521
D 803 NRHDNRSDNNTGMVLSPLYLNTTVLPSSSSSKGSLDSRSEKDLERERIGLGN 862
QY 522 --LSETPSSMAEKKVASLPO--DYDEPN--KTCESSNTSATTTSI-----QPNLEN 570
D 863 HPATENPTGTSKRGIOITSTAAQIAKVEEYSAIHTSQEDRSSGSTLHCVTERNALR 922
QY 571 SNSSSELNSSOSEKAKADDPENGRESHRTVSTIOELVGDFTSEKSTGELSESGAGK 630
D 923 RSSAAHSHNNTYNTKTS---ENSNRTCSMPYA---KLEYKRSSNDLSNYSSSDYGKR 975
QY 631 ASGSTRITRLRNPDSKLSQLKS--QCVAAAHEANKL-FKEGKEVLVYN----- 677
D 976 GOMKPSIESYSEDDSEKSCSGOYPRADLAHKIHGANHMDNDGELDPIINSLKYSDEQL 1035
QY 678 SOGELSLSTRKEVIMKGINNRYKLOEGRYRYHNOYSTNSPALNKHQREDHK-RR 736
D 1036 NSGRQSPQNERMARPKHIIIDEIKOSRQSR--NOSTT--YPVYTESYDDKHLKQF 1090
QY 737 HLAHFCILTAGEFKKNQSVHSGKVLITSLRLITIOLENNIPSEFFHPNASHRANKIK 796
D 1091 HFGQOECSVPYRSGANGS-----ETNRVGS-----NGLINQNV 1124
QY 797 AVOMCSKPRREFALALALIECAVRYVMLPIWREFLGRHLRHMTSIEEKEKVKKEKK 856
D 1125 SQSLCOED-----DIEDOKPTNYSERY 1146
QY 857 QEEETMOQATWVKYTPPVKHQVKKQGEERYVTGSGWMSIKTHVYRPVPLKPGNTNV 916
D 1147 SEEE-----OHEEERPTNY-----SI 1163
QY 917 NYKSLBGTKNMNDENDES-----DKKCSPPKIKIEPDESEKDEV 959
D 1164 KYNEE-----KHVDPQIDYSLKATDIPSSOKOSFSSKSSSGSSTKEHSSSEMTST 1219
QY 960 KGSDAANGADONEMDISKITEKKDQVKELDSDS-----DKP-CKEEMEV 1005
D 1220 PSSNAKR---QNLHPSSAQRSGQPOKAATCKKSSINQETIQYCVEDPIGCSRSSSL 1276
QY 1006 DDDMKTESHVNGQESSQ-----VDVYVNSEGFHLRT-----SYKKTKSKSL 1047
D 1277 SSLSSADELIGCNOTQOADSANTLOJLAIEKEKIGTRSADPDVEYPAVSQHPRTKSRSL 1336
QY 1048 DGLLEKRIKQFTLEKQRLKLEIKIGKIGKITSNSKN-----LSESVIT-TAKKEG 1101
D 1337 QG-----SLSSESARHNAVEPSSGAKSPSKGAQTTPKSPREHYVOETPLMSSRCTSY 1389
QY 1102 COSPMOEOBSPNANNPOPEDLLIOGCSODSSVYLRMSDPSHTTKLTKPKRVLDDVIRS 1161
D 1390 SSLDSFESRSIASSVQSEP-----CSGMVSGIISPSDLPDSPEQIIMPSPR----- 1434

QY 1162 PETKCPKONSLENDIEEKVSDLASRQEPTRKTKTGNDFFIDSKLASAD----- 1211
D 1435 SKTPEPPPTQATQTR-----VPKNKAPTAKEKRESG-----PKQAAVNAAGRVQVLP 1482
QY 1212 DIGILLI-----CKKK-----KPLIOESDITV-----SSKSKALHSSVPK 1246
D 1483 DADTLLEHATESTPDGFCSSLSALSILDEPFIQKDVRLIMPVQENDNOMETESQPK 1542
QY 1247 STNDRDATPLSRAMDFESKIGCDESENSTLENSSDTVSIODSPEDMIYONSNSISEQF 1306
D 1543 ESNENQ-----EKEAETKIDSEQ-----LDDSDDD----- 1569
QY 1307 RTRODVEVLEPLKCELVSGESTNCEDRLPVKSTEANGKSPQOKKLEERPVKCSDOI 1366
D 1570 -----DIEILE-BC-IISA-----MPTKSSR-KAKPAQTASKLPPVAKRKSQ 1611
QY 1367 ---KLKNTDDKKNNEENRESEKKGQRTSTFQINGDNKPKIYLKGECLKELSESRYVSGNV 1423
D 1612 PVYKLLPSQNRLOPKHVSFTPG-----DMPRVY---CY---EGTPINFT 1652
QY 1424 EPKVNNINKIIPENDIKSLVYKESAIRPFINGDVIMEDFNERNSSEKSHLLSSDAEGN 1483
D 1653 ATSLSDLTIESPPMEL-----AAGEVYAGQSGEFKEKDTIPEGR--STDEAQGG 1702
QY 1484 YRDSLETLPTSKES-----DSTQTTTPASCPES 1512
D 1703 KTSVY-TIPELDNKAERBDILACINSAMPKSKSHKPFYRYKIMDVOQASASSAPNK 1761
QY 1513 NSVNOVEDMEIETSEVKYVTS-----SPITSEESNLSNDF 1548
D 1762 NQLDGKR--KKPTVPKPIQONTERTYRTRVRKNADSKNNLAERYVSDNKDSKONLKNNS 1819
QY 1549 IDENGCLPINKENYNGE-----SKRKTY 1571
D 1820 KDFMDKLPNNEDRKRGSAFAPSPHHTYPIEGTPCFGRNSLSLDDDDVDLSRK-- 1877
QY 1572 ITFVNTMTSTVATESKTVIKVEKGQOTVYSTENCASKSTYT-----TTTITVTK 1621
D 1878 -AELRAKAKENESBAKVTSHTELSNQASANKQAIKQPINRQOPKPILOKOSTPQSS 1936
QY 1622 LSTPSTGSDVDIIVKQOSKTVYTTVYTDLSITTTGGGLVTSMYSKYSTROKVKLMKFS 1681
D 1937 KDIPDRGAAND--EKLQNALEIENTPVCFs--HNSLSLSLSDIOENNNKENEPIKETE 1990
QY 1682 RP-----KTRRGTAALPSYRK-----FVTKSTKSTIVLPDDE-----K 1716
D 1991 PPDSQGEPSKRPQASGYAPKSFHVEDTPVCFSRNSLSLSLSDIEDLLQECISAMPKK 2050
QY 1717 KLARKGIREVYPYFNNAKP-----ALDIWYPSRPPTFGIT-----WRYRL 1758
D 2051 KPSRLKGDNE-----KHSPRNMGGILGEDLTLDKIQIPDSSEHGLSPSENMWKAIQ 2104
QY 1759 QTVKSLAGVSLMLRLMASLRMDMAKVPGGSGSTSTSHSEHITITTELIKRRDQVPG 1818
D 2105 EGANSIV-----SLHQAATAACL-----SKQASDSD--SLSLKS--G 2140
QY 1819 IRFEYCIRKIICPIGVEPETKETPTPOKGLRSALPPKRPETPKQGPVLIETWVAEE 1878
D 2141 I-----SIGSGPHLLTPDDEKPFSTNKKPR--ILKPGKST-----LETKLIESE 2183
QY 1879 -----LEMEIRAFARVEREKEKAQAVEQAKRLDQKP-----TVIATST 1919
D 2184 SKGIGKGKVVYKSLITGKVRNSB-----ISGQMKPQLANMPSISRGMTMHIHP 2234
QY 1920 TSPFSTSTIISPAOKYMAPIGSAVTGTGMVLTITKVSFPAVITPOQKNFHQTAIYW 1979
D 2235 VRNSSSTSPSVSKGPELTKPASKSPBEGQTATTPSGAKPVS----- 2277
QY 1980 KOGSNGGVYVOQKVLGIIIPSSGTGSOQFTSFQOPTAVTIRMPNMGSGGTTMSQVY 2039
D 2278 ---KSELSPVARQTSQIG-----GSK-----APSRSGSRDSTPS----- 2309


```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match      2.4%; Score 362.5; DB 4; Length 2842;
Best Local Similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

Oy 373 WEAECKLIEEEREETIHRIMDTEDLTNRKSGNSKSFLLAANEELISIRAKKGIDIDNYK 432
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 698 WDMGAVSMKKNLIHKKHMI-----AMGS-----AALRLNLMANPAKTKDA-NIM 742

Oy 433 SPEET-----EKDN-ETENDSKDAEKNREEFEDOSLEKSDDDKTPDDPEOGKSEPT 485
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 743 SPGSSLPSTLHVAKOKALELDAQHLSETFDINDLSPKASHRSKQRIKOSLYGDVFPDT 802

Oy 486 EYGDGKNSVSAALGD-----NTT-----NAISEETSPBEGSGPVC----- 521
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 NRHDNRSDNFTGMNTVLSPLANTVLPSSSSSGSLDSSSEKDRSLERERIGLGNV 862

Oy 522 --LSETPSSNAEKKVASELPQ--DVPEEPN--KTCESSNTSATTTSI-----QPNLEN 570
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 HPATENPGTSSKRGIOISTTAQIAKVMEEVAIHTSQDRSSGTTTELHCYTDENALR 922

Oy 571 SNSSSELNDSOGESAKAADPENGERESHTPVSIQEIYVGFTEKSTGSELSESPGAGK 630
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 922 RSSAHTHSNTNYFTKS-----ENSNTCSMPYA--KLEYKRSSNDLSNVSSSDGYKR 975

Oy 631 ASGSRITRLRNPDSKLSQLKS--QVAAAHAEANKL-FKGEKVLVYN----- 677
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 976 GOMKPSIESYSEDDSGKECSYQOYPAADLAHKIHSANHMDDNGELDPTPIINYSLKYSDEQL 1035

Oy 678 SOGEISRLSTKKEVIMKGINNYFKLOGKRYVYNQYSTNSFALNKQHREPDHK-BR 736
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1036 NSGRQSPQSNEMWARKPIIIEDEIKOSEQRQR--NOSTT--YPYTTESTDDKHLKFPQ 1090

Oy 737 HLAHFCFLTPAGEFKWNGSVHSGKVLITSLRLITTLQLENNIPSSFHPNMAHSHRANWK 796
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 HFGQOECEVSPYSRGANGS-----ETNRGSG-----NHGINQAV 1124

Oy 797 AVOMKSKPREFALALALECAVAPVYMLPIKREFLGHTLHMTSLIEREKEKVKKKEK 856
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1125 SOSLQOED-----DYEDDKPTNYSERY 1146

Oy 857 GQEEETMOATWVXYTPVKHQVMKQKEEYRVYTGSGWSIKTHVYFVFKLPENTAV 916
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 SEEE-----QHHEERPTNY-----SI 1163

Oy 917 NYRKSLEGTKNMMDENMDS-----DKRKSRSPPKIKIEPDSKEDEV 959
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1164 KYNEE-----KRHVDPIDYSLKYATDIPSSOKOSFSSKSSSGGSKTTHMSSSSNTST 1219

Oy 960 KGSDAKAGDQWEMDISKTEKKDDVKLLDSDS-----DKP-CKEPEMEV 1005
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1220 PSSNAKR--QMLHPSSAQSSRGOPQKATCKVSSINOETIOTYCEVEDTPICFSCSSL 1276

Oy 1006 DDDMTESHVNOECSQ-----VDVNVVSGFHLRT-----SYKKTKTSKTL 1047
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1277 SSLSSAEDEIGCNOTTQEDASANTLQIAIEIKKIGTRSADEVSEVPAYVQHPRTKSSRL 1336

```

```

Oy 1048 DGLERRIKQFLEEKORLEKIKLEGKIGKGTSTNSSKN-----LSSSPVI-TKAKEG 1101
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1337 QG-----SSLSESAARRKAAVEFSSGAKSPSKSQGTQPKSPPEHYVQETPLMFSTCTSV 1389

Oy 1102 CQSDSMKQOSPPANNDQPEDLIQGCOSODSVLRMSDPSHTNKKIYPPDRVLDDVSISS 1161
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1390 SSLDSFESRSIASSVQSEF-----CSGAVSGIISPSDLDPDSGQTMPPSR----- 1434

Oy 1162 PETKCPKQNSIENDIEEKYSDLASRGOEPTKSKTKGNDFFIDSKLASAD----- 1211
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1435 SKTPPPPTQATQKRE-----VPKAKAPTAEKREG-----PKAAVANAQVQVQLP 1482

Oy 1212 DIGTLI-----CKNK-----KPLIQESDPTV-----SSSKALHSSVPK 1246
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1483 DADTLHFATETSPDGFSCSSLSLALSDLEPFOKDVLEIRIMPVQENDNGNETESEQPK 1542

Oy 1247 STNDRATPLSRAMDEGKLGCDSESNSTLSESDPVSIQDSSEEDMIYQNSNESTSEQF 1306
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1543 ESNENQ-----EKEAKTIDSEK--LILDSDDD----- 1569

Oy 1307 RTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTAEANKKPPSOOKLEERPYNKCSDOI 1366
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1570 ----DIEILE--EC-IISA-----MPTKSSR-KAKKPAQTASKLPPYARKRPQL 1611

Oy 1367 ---KIKNTDKKNENRESEKKQGTSTFOINGKDNKPKIYLKCECLKETISESRVSGNV 1423
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1612 PVYKLLPSQNRLOPKVHVSFTPG-----DDMPRYV--CV-----EGTPIINFST 1652

Oy 1424 EPKVNINKIIPENDIKSLTVKESAIRPFIINGVIMEDENERSSTFKSHLLSSDAEN 1483
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1653 ATSLDLTIESPPNEL-----AAGEVGGAQSGEPEKRTJITEGR--STDAQGG 1702

Oy 1484 YRDLSTLPTSKES-----DSTQTTTPSPASCPE 1512
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1703 KTSVV-TIEBLDNKKAEBGDILAEICINSAMPKGSKPPRVKAKIMQVQOASASSAPNK 1761

Oy 1513 NSVNOYEDMEIETSEVKKYTS-----SPTSEESNLNDF 1548
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1762 NOLDGKRC--KKPTSPVKPIQNTERTVRKNAKSNLNLAERVFSDNKSQONKKNNS 1819

Oy 1549 IDENGLPIKNENVNE-----SKRKTIV 1571
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1820 KQFNDRKLPNNEDEVKGFAPDSPHHTPIEGTPIYCFSRNDSLSLDFDDDDVDLSREK-- 1877

Oy 1572 ITEVTMTSTVATSEKTVLKEKDKQTVVSSPENCAKSTVT-----TTTTPVTK 1621
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1878 -AELRAKENKESEAVYTSHTELTSNOQSANKQALAKOPINNGQPKPIIQKSTPQSS 1936

Oy 1622 LSTPSTGSAVDIISVKEQSKTVVTTVDTSLTTGGTLVTSMTVSKEYSTRDKVLMKFS 1681
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1937 KDIPDRGAATD--EKLQNPATIENTVCFSS--HNSLSLSLSDIDQENNNKENEPIKETE 1990

Oy 1682 RP-----KTRRSALSYRK-----FYTKSTKSIYFLPNDUL--K 1716
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1991 PPDQGEPEKPPQASGYAPSFHVEDTPVCFSRNSSLSSLSISEDDLQECISSAMPKK 2050

Oy 1717 KLAARKGIREVYFYNNAKP-----ALDIPIYSPRPFTGFT-----WRYRL 1758
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2051 KPSRLKGDNE-----KHSFRNMGILGEDLTLDLDIQDRSEHGLSPDSENFPMKALQ 2104

Oy 1759 QTVKSLAGVSLMLRLMLASLRMDMAKAVPPGGSGTRETSETETITTEITIKRDVGYG 1818
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2105 EGANSIV-----SLHQAAAAACL-----SRQASSDSD--SILSKS---G 2140

Oy 1819 IREYCIKIKIICIGVPEIPKERTPQKGLRSALRPKRPETPKOTGVPIITWAAEE 1878
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2141 I-----SLGSPFHLTPDOEEKPFTSNKGP--ILKPEKST-----LETKRIESE 2183

Oy 1879 -----LELMETRAFAERVEKEKAQAVEQAKRLLEQOKP-----TVYATST 1919
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2184 SKGIGKGRKVKYSLITGKRSNE-----ISQMKQPLQANPPSISRGTMTHHG 2234

Oy 1920 TSPTSTSTSTISPAOKVWVAPISSGYTGTGMVLTTKVGSPPATVTPQONKNFHQTFATV 1979

```



```

Db 2235 VRNSSSTSPVSKKPPPLTPAKSKSPSECOATATTPRGAKPSV----- 2277
QY 1980 KOGOSNGVVOQVKVLIIPSSSTGTSGOTFTSFOPRTAVTIRPTVSGSGGTTNSOVL 2039
Db 2278 ---KSELSVPARQTSOIG-----GSSK-----APSRSGSRDSTPS----- 2309
QY 2040 TGPQIRPGMTVIRTPLOQSTLGAITRTPYMVOPG---APQVMTQIIRGQPVSTAVSA 2095
Db 2310 ----RPAQOPLSRPIQSPGRNS-----ISPGRNGISPPKRLSOLPRTSSPTA--- 2353
QY 2096 PNTVSSPFGOKSLTS-----ATSTNIOSSASOPPPQOGVKLMAOLTQTLQGH 2146
Db 2354 -STKSSGSGMSTSTPGRQMSQONLKQICLSKNASSIPRSESAS-----KGLNQNNGV 2407
QY 2147 GGNQGLTVLIQGGQTTGOLQ-----LIPQGVLPFGGQLOLMQAMPNGTVQRFLE 2199
Db 2408 GANKKVELSHMSSTKSSGSDRSERPVLRQSTFIKEAPSPILRRKLEESAS-----FE 2462
QY 2200 PLATTTATTTTSTTSTTACGTGEBQROKLSLSP-----QMQVHQD-----KTLPPAQS 2248
Db 2463 SLSPSSRPAPSTKSOAQTPV-----LSPLPDMSLSTHSSVOAGGWRKLPPLSP 2512
QY 2249 SVGPAAQOPOTAPQAPQOPOTOPQSPAPQEPVOTOTTVSSHVSEAPHTAQS 2308
Db 2513 TIEYNQRPAPKRDIAHSHSESPRLPINSGTWKRE---HSKHSSLPVSTWRRRTGSS 2569
QY 2309 KPOVAASQPOQSVNOGQSPYRVQSPQOTRIRPTPSOLS-PCGOQSOVQITTSQPIPIQPH 2367
Db 2570 SSILSSSESESEKAKSDEKHVNSISGT--KQSKENOVSAKGTWRKLIKENEFSF----- 2621
QY 2368 TSLQIPSGQPOQPOQSVOSTOTLSSGOT 2396
Db 2622 -----TNSTSQTVSSGAT 2634

```

```

RESULT 10
; Sequence 2, Application US/08452655B
; Patient No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-655B-2

```

```

Query Match 2.4%; Score 362.5; DB 1; Length 2843;
Best Local Similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

```



```

QY 1048 DGLLEERRIKQFLEEKQLEKIKLEGGIGICKTSTNSSKN-----LSSESPVI--TKAKEG 1101
Db 1338 QO-----SSLSESEARHKAVEFSSGAKSPSGAQOTPKSPENHYQETPLMFSTCTSV 1390
QY 1102 CQSDSMRQOSQNNANNNDQEDLLQGCSDSDSVLRMSDPSSHNTNKLYPXDRLVDVYIS 1161
Db 1391 SSLDESSESSIASVOSE-----CSGVSGIISPSDLPDSCQIMPPSR----- 1435
QY 1162 PETKCPKQNSINDIEEKVSDLASRQOEPTKTKGNDFIDSKLASAD----- 1211
Db 1436 SKTPEPPPTAQTKRE-----VPKKKATFAEKRESG-----PKQAAVNAVQRYVLP 1483
QY 1212 DIGTLI-----CKNK-----KPLQESDTIV-----SSKSALHSSVVK 1246
Db 1484 DADTLHFATFESTPDGFCSSSLALSLEDPFIQKDVLELRIMPVQENDNGNTESEQPK 1543
QY 1247 SUNDNATPLSRAMDEEGKLGCDSESNSTLENSDPTVSIQDSEEMIVQNSNESTSEQF 1306
Db 1544 ESENEO-----EKEAEKTISEKD-----LLDSDDD----- 1570
QY 1307 RTREDDVEVLEPLKCELVGSESTGNCEDRLPVKGTENGKSPSOQKLEBERPVNKCSDOI 1366
Db 1571 -----DIELLE--EC-IISA-----MPTKSSR-KAKKAQOTASKLPPVAKRPSOL 1612
QY 1367 ---KLKNTTDKKNNESESEKKGQRTSTFOINGKDNKPRIYLKGECLKEISESRVVSQNV 1423
Db 1613 PVYKILLPSQNRLOPKQHVAFSTPG-----DDMPRVY--CV-----ECTPIINFST 1653
QY 1424 EKVANNINKIIPENDIKSLTVKESAIRPINDVIMEDNERNRSTKSHLLSSDAEGN 1483
Db 1654 ATSLSDLTIESPNEI-----AAGEVKGAGSAGEFEKRDITPTER--STDQAOG 1703
QY 1484 YRDSLETLPTSKES-----DSTQTTTPSASCPS 1512
Db 1704 KTSVV-TIEELDDNKAEEEDIIAECINSMAPKGSHPKPRVAKINDQVOASASSAPNK 1762
QY 1513 NSVNOVEDMEIETSEVKVTS-----SPITSEBSNLNDP 1548
Db 1763 NOLDGKK--KKPTSPVKPIQMTXEYRTVRKRNADSKNNLNAERVFSDNKKONLKNNS 1820
QY 1549 IDENGLPIKKNNVNCE-----SKRKIV 1571
Db 1821 KQFNKPLPNNEBRVRSFSAFDSPHHTPIEGTPCYFSRNDLSLDDDDVDLREK-- 1878
QY 1572 IREVTMTSTVATESKTVIVKEGDKQIVVSSSTENCASKSTV-----TTTTVTK 1621
Db 1879 -AELKRAKNESEAKVYSHTELTISQOSANKTOAIKAPINRGOPKPILOKOSTPQSS 1937
QY 1622 LSTPSTGSGVDIISVKEOSKIVVTTVYDLSLTGGTLVTSMTVSKEYSTRDKVKLMKFS 1681
Db 1938 KDIPDGAATD---EKLQFAIENTPVCS---HNSLSLSDIDQENNNKEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FYKSKTKSTFVLPNDL-----K 1716
Db 1992 PPDSCGEPSPKQASGAPSPSFHVEDTPVCFSRNSSLISLSDSEDDLLQECISAMPKPK 2051
QY 1717 KLARKGJIREVYFYVNAKP-----ALDIMPSPRPRTGIT-----WRYRL 1758
Db 2052 KPSRLKGDNE-----KSPRMWGLGEDLTLDKIDIRPDSEHGLSDSENFDMKAIQ 2105
QY 1759 QTVKSLAGVSLMLRLMLMSLRMDMAKVPGGGSTRTETSETITTTTIIIRRDVPGY 1818
Db 2106 EGANSIV-----SLHQAAAAACL-----SRQSSDSD---SILSKS---G 2141
QY 1819 IRFECEIKIICPIGVPEPKPTTPQKGLSSALRPRPPTPKOTGVIIETVAAEE 1878
Db 2142 I-----SLGSPFLUTPDQEKPKPTSNKGP--ILKPEKST-----DETKEISE 2184
QY 1879 -----LELWEIRAFERKEKAQAVEQOAKKRLQOQP-----TVIATST 1919
Db 2185 SKGIGKGGKKVYSLITGKVRNSF-----ISGQMKQPIQANMPSISRGTMIHIG 2235

```

```

QY 1920 TSPSTSTTSTISPAQVYMAVIGSVYTTGKMLVLTTRKVGSPATVTOQKNPHOTATWV 1979
Db 2236 VRNSSSTSPVSKKGPPLTPAKSPSEGOATATSPGAKPSV----- 2278
QY 1980 KQOSNSGVVOYQKVLGIIPSSSTGTSQOTFTSFQRTATVTRPMTSGSGGTSMSOVI 2039
Db 2279 ---KSELSPAROTSQIG-----GSSK-----APSRSGSDSTPS--- 2310
QY 2040 TGPQIRGMATVITPLQOOSTLGKAIIRTPVMVOFG-----APOQVMQIIRGQPVSTAVSA 2095
Db 2311 -----RPAQOPLSRPIQSGRNS-----ISPGANGISPPMKLSQLPRTSSPSFA--- 2354
QY 2096 PNTVSTSPQKSLTS-----ATSTNSIQSSASQPPRPGQGVYKLTMAQLTQLTQGH 2146
Db 2355 -STKSSGSGKMSYSTSPGRMSQONLTKQTLGSKNASSIPRESAS-----KGLNOMNGN 2408
QY 2147 GGNQGLTVIYQGGQOTTGQLO-----LIPGVTVLPGGQOLMAAMPNGVQFFLT 2199
Db 2409 GAKKVELSRMSTKSSGSESDRSERPVLRQSTFIKEASPTRLRRLKEESAS-----FE 2463
QY 2200 PLATVATTAATTTTSTTAAGTGEORQSKLSP-----QMOVHOD-----KTLPPAQS 2248
Db 2464 SLSPSSRPASPTISOQOTV-----LSPLDMSLSTHSSVQAGQKRLPPLNLS 2513
QY 2249 SVGPANAQOTAPASARPOPOTQOPQSPAQPEVOTQOTVTSVHPSEAOPTHAQSS 2308
Db 2514 TIEYNDGRPAKRHDIAIRSHSESPRLPINSGTWKRE---HSKHSLSLPIVSTWRRTSS 2570
QY 2309 KPOVAAOSQOSVQOSQSPRVQSPQTRIRPSTPQOLS--PGQSOVQOTTSPITQIPH 2367
Db 2571 SSISSASSESEKAKSEDEKHVNSISGT---KQKRENOVSAGKTWRKIKENEFSP----- 2622
QY 2368 TSLQIPQOQPOQSOPOVOSTOTLSSGQF 2396
Db 2623 -----TNSTSQTVSSGAT 2635

```

```

RESULT 11
US-08-452-655B-7
; Sequence 7, Application us/08452655B
; Patent No. 5783666
;
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLIN, GEOF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
;
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
;
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
;
; NUMBER OF SEQUENCES: 102
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548

```


FILED DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ANTI-SENSE: NO
US-08-452-655B-7

Query Match 2.4% Score 362.5; DB 1; Length 2843;

Best Local Similarity 17.3%; Pred. No. 8.1e-13; Indels 777; Gaps 98;

Matches 409; Conservative 327; Mismatches 856;

QY 373 WEALCKILEMEBEIHRMDITEDLTNRKAGSNKSFLLAANEIIESIRAKKGDIDNVK 432
DB 699 WDMGAVSMLKNLHSHKHKMI-----AMGS-----AALNLNMANRAKTKDA-NIM 743
QY 433 SPREI-----EKDKN-ETENDSKDAEKNEEEDOSLEKSDDKIPDDPEOGKSEPT 485
DB 744 SPSSSLPSLHVROKALELDAQHLETFDNIIDNLSPKASHRSKORHKOSLGDVYVDT 803
QY 486 EVGDKGNSVANLGD-----NTT-----NATSEETSPSGRSPVGC--- 521
DB 804 NRHDDNRDNFNTGNTMTVLSPLYNTIYVLPSSSSSGSLDSRSREKDRSLEREGIGLGN 863
QY 522 --LSETPDSSNMAEKVASELPQ--DVPEEPN--KTCESSNTSATTTST-----QPNLEN 570
DB 864 HPATEPNTGTSKRGLQISTTAQIAKMEVSAIHTSQEDRSSTGSLTCLHCYDERNALR 923
QY 571 SNSSSELNSQSASAKAADPENGGERESHPTVSIQEIYVDFTSEKSTGLSSPGAGK 630
DB 924 RSSAAHTHSTTYFTKS---ENSNRTCSMPYA---KLEYKRSSNDLSVSSSDYGR 976
QY 631 ASGSTRIITRLRNPDSKLSQLSK--QOVAAAHEANKL-FKEGEVLYVN----- 677
DB 977 GQMKPSIESYSEDESKFCFSYGYQYPADLAKHISANHMDDNCGELDTPINYSIKYSEOL 1036
QY 678 SQGEIRLSTKKEVIMKGINNTFKLGQEGKYRVYHNOYSTNSFALNKHQHREDHDK-RR 736
DB 1037 NSGRQSPSONERWARPKHIEIEIKOSQROSR--NOSTT--YPYTESTDCHKHLKQP 1091
QY 737 HLAHKFCILPAGFEKNGSVHSGKVLITSLRLITQLNNIPSPFFHNMASHRANWLK 796
DB 1092 HEGQCEVSPYRKGANGS-----ETNRVGS-----NHGINOV 1125
QY 797 AVOMCSKPREFALALAILCAVAPVMLPIWREFLGHTRLHRMTSIEREKEKVKKEREK 856
DB 1126 SGLSCQED-----DYDDDKPTNYSERY 1147
QY 857 QEEEFEMOATVWKTTFPVKHOVMKOGEEYRVYTGSGWSIKTHVYRFVPLKPNITNV 916
DB 1148 SEEE-----OHEEERPTNY-----SI 1164
QY 917 NYRKSLEGTKNNDENMDES-----DKRKSRSPPKIKIIEPSEKDEV 959
DB 1165 KYNB-----KRHYDDPITSLKYAIDIPSSQKOSFSSKSSGSSKTEHMSSSSNTST 1220
QY 960 KSSDAKAGADQEMDISKITEKKDQVKEILDSDS-----DKP--CKEEMPEV 1005
DB 1221 PSSNAKR---QNLHPSSASQSRSGDPQKAATCKVSSINQETIQTVCVEDEPTLCFSRCSTL 1277

QY 1006 DDDMKTESHVNCQESSQ-----VDVYVNSEGFHLRT-----SYKKTSSKL 1047
DB 1278 SSUSSADEIGCQNTQOADSANTLQJAEIKETIGTASADPVSEVPAYSQHPRKSSRL 1337
QY 1048 DGLERRIKQFTLEEKORLEKIKLEGIKIGTSTNSKN-----LSESPVI--TKAREG 1101
DB 1338 QG-----SSLSSEGRHKAVEFSSGAKSPSKSGMOTPKSPPEHNVQETPLMFSTCTSV 1390
QY 1102 QSDSMKROESPPANNDQPEDDLIOGQSQSDSVLRMSDPSHTTNKLYPKDRVLDVYSRS 1161
DB 1391 SSLDSEFSRSIASSVOSEP-----CSGWVSGITISDDLDPDSGQIMPSPSR----- 1435
QY 1162 PETCKPKQNSIENDIEKVSADLSRQGPETKSTKNDPFIIDSKLASD----- 1211
DB 1436 SKTPPPPPQJQTKRE-----VPKNKAPAEKRESG-----PKOAAVAAVORVOYLP 1483
QY 1212 DIGTLI-----CKNK-----KPLIOESDITV-----SSKSALHSSVPR 1246
DB 1484 DADTLHLFATESTPDGFCSSSLALSJLDEPFIQKQVLELRIMPVQENDNGNTESEQPK 1543
QY 1247 STNDRATPLSRAMDEBGKIGCDSSENSSTIENSSTVSIQDSSEEDMIYONSNESTSEOF 1306
DB 1544 ESNEMO-----EKEAKTIDSEKD-----LIDSDDD----- 1570
QY 1307 RTREODVEVLEPLKCELVSGESTGNCEDRLPVKGTANGKPKSQOKLEBRPVNKCSDOI 1366
DB 1571 -----DIEILE-EC-IISA-----MPTKSSR-KAKKPAQOTASKLPPYARAKPSOI 1612
QY 1367 --KLKNTYDKNNENRESEKKQRTSTFOINGDKNPKIYKGECLKEISESRVYSGNV 1423
DB 1613 PVYKLLPSQNRLOPQKHVSFTPG-----DOMPRVY-----EGTPIINFST 1653
QY 1424 EPKVNINIKLIPENDIKSLVKSALRPFINGVIMEDENERSSTKSHLSSDAEN 1483
DB 1654 ATSLSDLTIESPNEI-----AAGEVGGAQSGEEFKRTIPTER--STDAQOG 1703
QY 1484 YRDLSTLPSTKES-----DSTQTTSPASCPES 1512
DB 1704 KTSVV--TIPELDNKAEBGDIACINSAMPKSHKPPRVKIMQYQOASASSAPRK 1762
QY 1513 NSVNOYEDMEIETSEVKKYTS-----SPITSEBSNLSNF 1548
DB 1763 NOLDGKR--KKPTSPVKKPIPONTERTVRKNKADSKNLNAERVSNDKSKRONKNS 1820
QY 1549 IDENGLPIKKNENVNGE-----SKRKTV 1571
DB 1821 KDFNDKLPNNEDRVKSFAPDSPHHTPIEGTPYCFSRNDSLSLDPDDDDVDLSREK-- 1878
QY 1572 ITEVTMTSTVATESKTVIKVKGDKQTVVSSSTENCAKSTVT-----TTTTVYTK 1621
DB 1879 -AELRAKENKSEAKVHTSHTEITSNQSANQTQALAKQPIRNGQPKPILQKQSTFPQSS 1937
QY 1622 LSTPTSGSVDIISVKEQSKYVTTTIVDTLTTGGTIVTSMTVSKESVTRKQVLMKFS 1681
DB 1938 KQIPDGAATD--EKLQFALENTPVCS--HNSLSLSLDIOENNKNKENEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FVTKSTKSKSFVLPNDL-----K 1716
DB 1992 PPDGSGEPKPAQSGAARPSFHVEDTPVCFSRNKSLSLSIASEDDLLOECISSAMPKK 2051
QY 1717 KLARKGGIREVPYFVYMAKP-----ALDIWPPSPRPRTFGIT-----WRYRL 1758
DB 2052 KPSRLKGDNE-----KHSFRNWGILGBDLTLDKIDQRPDSEHGLSPDSENFQWAKIQ 2105
QY 1759 QTVKSLAGVSLMLRLMLASLRMDMAAKVPPGSGSTRTETSTETLTTEIIRKRDVGYG 1818
DB 2106 EGANSIV-----SLHQAANAACL-----SROASSDSD-----SILSKS--G 2141
QY 1819 IFFECIRKIIQPIGVPEPKETPTPORKGLSALARKRPRTPKQGTGPVILEETWVAEE 1878
DB 2142 I-----SLGSPFHLTPDQEKPFSTSNKGR--ILKFGKST-----LETKIESE 2184


```

QY 1879 -----LELMETRAFAERVEKEKAQVEQAKKRLQOKP-----TVIATST 1919
Db 2185 SKGKGGKVKVYSLTGKRSNSE-----ISGOMKOPLOANPMSISRGRTMHIHG 2235
QY 1920 TSPSTSTSTISPAQKVMVAPIGSVTGTGMVLTTKVGSFATVTPQKNFQOTATW 1979
Db 2236 VRRSSSTSPVSKGPKPLTPASKSPSEGOATTSRGAKPSV-----2278
QY 1980 KQGQNSGVVQVQKVLGIIIPSTGTSQOTFTSPORATVTRPNTSGSGGTTNSQV 2039
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APRRSGSRDSTPS---- 2310
QY 2040 TGPQIRPGMTVTRPLQOSTLKAITRTVMVOPG-----APQOVMTQIIRGQVSTAVSA 2095
Db 2311 -----RPAQOPISRPQIOSGRNS-----ISPGRNGISPPNKLQPLPTSSPSTA--- 2354
QY 2096 PNTVSTPGOKSLTS-----ATSTNSIOSASQPPRQOGQVKLMAQLTQLQGH 2146
Db 2355 -STKSSGSKMSTTSPGRMSQONLTQGTGLSKNASSIPSESAS-----KGLNOMNNGN 2408
QY 2147 GGNQGLTVVIOGQGTGQLQ-----LIPQVTVLPGRQOLQOAMPNGTVQRFLEF 2199
Db 2409 GANKVELSRMSTKSGSESDRSPVLRQSTFIKAPSPTLRRKLEESAS-----FE 2463
QY 2200 PLATATATSTTTTSTTAAGTGEOROSKLS-----OMQVHOD-----KTLPPAQS 2248
Db 2464 SLSPSSRPSPRSQAQTV-----LSPLSDPMSLSTHSSVQAGWKRKLPLNLS 2513
QY 2249 SVGPAKAPQOTAPSAPOPOPTOPPOPAPOPEVOTQOTVSHVSEAPTHAQS 2308
Db 2514 TLEYNGRAKAKHDIARSHSESPSLPINSRGTWRE---HSHSSSLPVRVSWRTGSS 2370
QY 2309 KQOVAQSQPOQNSQGVVQSPQSPQTRIRPSTPQLS-PGQSQVQVTTQSPPIPIOPH 2367
Db 2571 SSILSSSESSSEKAKSEDEKHNVISIGT--KQSKENQVSAKGTWRIKENEFSF----- 2622
QY 2366 TSLQIPSQGQPOQPOVOSTQTLSGOT 2396
Db 2623 -----TNSTQTVSSGAT 2635

RESULT 12
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582

```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-582-2

Query Match 2.4%; Score 362.5; DB 3; Length 2843;
Best local Similarity 17.3%; Pred. No. 8,1e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

QY 373 WEDELKILEEMREELIHRHMDITEDLTNKAQSGNKSFLAANEELIESIRAKKGIDIDNYK 432
Db 699 WDMGAVSMKLNLIHSHKHM-----AMGS-----AAALRNLMANRPKAKYKDA-NIM 743
QY 433 SPEET-----EKDN-ETENDSKDAEKNEEFEDOSLEKDSODKTPDDDEQGXSEPT 485
Db 744 SPQSSLPPLHVRKQKALEALDQHLSETFDNTIDNLSPKASHRSKORHKOGLYGVYFDT 803
QY 486 EVGDKGNSVANLGD-----NTT-----NATSEETSPSEGRSPVC--- 521
Db 804 NRHDNRSDNFMNTGNMTVLSPLNTVLPSSSSRGLSDSSREKRSJLEREGIDLGNY 863
QY 522 --LSETPDSSNNAEKVYASELPQ--DYPEEPN--KTCESSNTSATTTST-----QNLN 570
Db 864 HPATENRGTSKRGLOISTTAQIAKVMEEVSAIHTSOEDRSSGSTTELCVYDENNAIR 923
QY 571 SNSSSELNSQSESAKAADDPENGESRHTPVYIQEIIQDFTSEKSTGELSPGAGK 630
Db 924 RSSAHTHSNTVFTKS-----ENSNRTCSMPYA--KLEYKRSNDLSNVSSSDGYGR 976
QY 631 ASGSTRIITRLRNPDSKLSOLKS--QOVAAAHAEANKL-FKEGKVLVYN----- 677
Db 977 GQWKPSIESYSEDESKFCYSYGYPADLAHKIHSANHMDDNGELDTPINYSIKYSDEOL 1036
QY 678 SOGEIIRLTKEKVKVINGKINNNTFKLGQEGKRVYINQYSTNSFALNKQHREDHDK-RR 736
Db 1037 NSGRQSPQSONERWARPKHIIIEIKOSEQORSR--NOSTT--YPVYTESTDCKHLKFP 1091
QY 737 HLAHKCLTPAGFEKNGSVHSGKVLTLSTLRITLQLENNIPSPFHHNMASHRANWIK 796
Db 1092 HFGQDECVSPYSRGANGS-----ETNRVGS-----NHGINOV 1125
QY 797 AVQMSKPREFALAIALILECAVAPVYMLPIWREFLGHTPLRHMTSIEREKERVKKKK 856
Db 1126 SOSLQGED-----DYEDDKPTNVSEHY 1147
QY 857 QEEELMQOATVTKTFPVKHQVWKQGEERYTYGSGWSISKTHVYFVFKLPGNTNV 916
Db 1148 SEEE-----OHEEBEPTNY-----SI 1164
QY 917 NYRKSLEGTKNMNDENWDES-----DKRCSRPSPKIKIIPDSEKDEV 959
Db 1165 KYNEE---KRHYDPIIDYSLKAYDAIDIPSSQKQSFSSKSSSGQSKTBHMSSESTNT 1220

```



```

QY 960 KGSDAKAGADQONEMDISKITEKKDQVKELLSDS-----DKP-CKEEMEV 1005
Db 1221 PSSNAR---QONLHPSAQSBSGQPOKATCVCSSINQETIQTYCEVEDPIFCSSJ 1277
QY 1006 DDDMKTESHVNCOESQ-----VDVNVSEGFHLRT-----SYKKKTSKL 1047
Db 1278 SLSSEDEDELGCNOTQOEDSANTLQIAETKEKIGTRSAEDVPSEVPAVSQHRRTKSRL 1337
QY 1048 DGLLEIRIKOFTLEEKORLEKIKLEGKIGKIGSTNSSKN-----LSESPV-TRAKEG 1101
Db 1338 QG-----SLSSESARHKAVEFFSSGAKSPKSGAOTPKSPENHYOETPLMFSCTSV 1390
QY 1102 COSDSMKROEOPNANDQPEDLIQCSQSDSVLRMSDPSHTNKLKPKDRLVDVIRS 1161
Db 1391 SLSDESERSIASSVOSEP-----CSGWYSGLIISPDLPDPSGQIMPPSR----- 1435
QY 1162 PETCKRKONSTENDIEKYSDLASRGOEPTKSTKGNDFIDSKLASAD----- 1211
Db 1436 SKTPPPPOQTAKTRE-----VPKNKAPTAEKREG-----PKOAAVAAVQROVLP 1483
QY 1212 DIGTLI-----CKNK-----KPLIOESDITV-----SSSKALHSSVPK 1246
Db 1484 DATLLHFAFESTPDGFCSSLSALSDEPFIQKDELRLMPVOENGNENTESBQPK 1543
QY 1247 STNDRDATPLSRAMDEEGKLGCDSESNSTLENSDVTYSIDSSSEDMIVONSSESIEOP 1306
Db 1544 ESNENO-----EKEAKETIDSEKD-----LLDSDSD----- 1570
QY 1307 RTEQOVEVLEPIKCELVSESTGNCEDRLPVGTEANGKPKSQOKLEPRPNKCSOI 1366
Db 1571 -----DIEILE-EC-IISA-----MPTSSSR-KAKPAQOTASKLPPPAKRPQSOL 1612
QY 1367 ---KLKATTKKNNENRESKKGORTSTFQINGKNKPKTYLKEGCELTSESRVSGNV 1423
Db 1613 PVKKLPSQRLOPQKHVSTFG-----DDMKRY-----CV-----EGTPIINFST 1653
QY 1424 EPYVNNINKIIPENDIKSLTVKESAIRPPIJNDVIMEDENERSSETKSHLLSSDAEGN 1483
Db 1654 ATSLSDTLTISPPNEL-----AAGEVGVGAQSGEPEKRDITPEGR--STDEQOG 1703
QY 1484 YRDSLETLPSTKES-----DSTOTTPSASCPES 1512
Db 1704 KTSGV--TIPELDKNKAEGDILAECTINSAMPKGSKHPFKKIMDVQOASASSAPNK 1762
QY 1513 NSVNOVEDMEIETSEVKYVS-----SEPTSEESNLSNDF 1548
Db 1763 NQDQKK--KKPTSPYKPIQNTETRYRVRKKNADSKNNLNAEVEFSNKNKSKQNLKNS 1820
QY 1549 IDENGLPIKNENVNGE-----SKRKIV 1571
Db 1821 KDENDKLPNNEEDVRGSFAFDSPHHTPIEGTPEFSRNDLSLDFDDDDVLSREK-- 1878
QY 1572 ITEVTMTSTVATESKTVIAVEKGDKQTVVSTENCACKSTVT-----TTTTVTVK 1621
Db 1879 -AEIRAKAKEKESAKYTTSTELTSNQOSANKQOAIKQPIINGQPKPIIKQKSTFPOSS 1937
QY 1622 LSTPSTGGSVDIISVKEQSKTVYTTVTDLSITTGGLVTSMTVSKESYSTRDKVLMKFS 1681
Db 1938 KDIPDRCAAND---EKLONPAIENTPVCFS---HNSLSLSIDIDQNNKKNEMPIKETE 1991
QY 1682 RP-----KTRSGTALBETRK-----FVTSTKKSITVLNDDL-----K 1716
Db 1992 PPSQSGPSPQASGYAPKSFHVEDTPVCFSRNSSLSSLSIDSEDDLOECISSAMKPK 2051
QY 1717 KLARKGIRFVPEFVNAKP-----ALDIMYPSPRPPTFGIT-----WRYL 1758
Db 2052 KPSRLKGDNE-----KHSRNMGGILGEDLTLDIKDIQRPDSEHGISPSSEMPDKAIQ 2105
QY 1759 QTVKSLAGVSLMLRLMASLRMDMAKVPBGGSSTRTETSETEITTEITIKRDVGPYG 1818
Db 2106 EGANSIV-----SSLHQAAAAACL-----SRQASDSD--SILSLKS---G 2141
QY 1819 IREYCIKRIICPIGVETPKETPTPQKGLBSALRPKRPETPKQGPVITETWAEEL 1878

```

```

Db 2142 I-----SIGSEPHLTPQOEKPEFTSNKGP--ILKPEKST-----LETKIESE 2184
QY 1879 -----LEMEIRAFAREYKEKEQAQAVEQAKRLQEQK-----TVIATST 1919
Db 2185 SKGIRKGGKVVYKSLITGKRSNE-----ISQOMKOPLODAMPISIRGTMHIHG 2235
QY 1920 TSPSTSTTISPAQWMAPISGSVTTGKVLITTKVSPATVTPQOKNKEHQEPATWV 1979
Db 2236 VRNSSSTSPVSKGKGPPLATPAKSPSEGGTATTPRGAKPV----- 2278
QY 1980 KQOSNSGVYVOQKVLGIIIPSTGSOOTFTSFQRTATVTIRPMTSGSGGTTNSQVI 2039
Db 2279 ---KSELSPYARQTSQIG-----GSSK-----APSRGSRDSTPS----- 2310
QY 2040 TGPQIRPGMTVIRTPLOQSTFLGKAILRTPVWOPG---APQOVMTQIIRGQPVSTAVSA 2095
Db 2311 ---RPAQOPLSRPIQSPGRNS-----ISPGRNGISPRKLSQLRSTSPSTA--- 2354
QY 2096 PNTVSTPGQKSLTS-----ATSTSNIOSSASQPPRPOGOYKLTMAOLTQULTQGH 2146
Db 2355 -STKSSGSGKMYTSPGRQMSQONLTKOTGLSKNASSIPRESAS-----KGLNQNNNG 2408
QY 2147 GGNQGLTVYIQQGGQTTGQLQ-----LIPQGVYLPBPGQOLMAAMPNGVQFLFT 2199
Db 2409 GANKKVELSRMSTKSSGSESDRSEPRVIVRQSTFTKEASPTLRKRLLESAS-----FE 2463
QY 2200 PLATTTATTAATTTTSTVTAAGTGEOBKSLP-----OMQVHQD-----KTLPAOSS 2248
Db 2464 SLSPSRPASPTISOQATPV-----LSLSLDMSLSTHSSVQAGWRKLPPLNLSR 2513
QY 2249 SVGPAAQOPQTAOPASRPQOTQPOSPADQEVQTOPEVQOTQTVSSHVSEAOPTHAOSS 2308
Db 2514 TIEYNDGRPAKRHDIAHSHSESPRLINRSQGWKRE---HSKHSSSLPRVSTWRRTGS 2570
QY 2309 KPOVAAOSOPQSVQOQSPVRYVQSPQTRIRPTSPOLS-PPGOQVOYTTISQPIPIQH 2367
Db 2571 SSILSASSESEKAKSEDEKHNVISGT--KQSKENOVSAKGTWIKIKENESP----- 2622
QY 2368 TSLQIPSGOPQOSQPOVOSTQTLSSQOT 2396
Db 2623 -----TNSTSQTVSSGAT 2635

RESULT 13
US-08-450-582-7
; Sequence 7, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSTLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-450-582-7

```


QY 1759 QYKSLAGVSLMLRLMASLRMDMAKVPGGSGTFTSTETITTEITTEIKRDVPGY 1818
Db 2106 EGANSTV-----SSLHQAAAAACL-----SFQASDSD-----STLSKS-----G 2141
QY 1819 IREYECIRKIIICPIGVETFKETPTPOKGLRSSALRPKPKETPKOTGPVIEFWAEE 1878
Db 2142 I-----SLGSPFHLIPDDEKPFISNKGPR--ILKPGKST-----LETKKESE 2184
QY 1879 -----LEMEITAPAEERKEKKAQAVEOQAKRLDQKF-----TVIATST 1919
Db 2185 SKGKGGKKYKSLITGKVSNSP-----ISGQMKOPLQANMPSISRGFTMHIPG 2235
QY 1920 TSPSSSTSTISPAQKMWAPISGSVTTGPKMVLTFVKGSPATVTFQONKNHQTFTW 1979
Db 2236 VRSSSTSTSVSKGPKPLTPAKSPSEGTATTSPGAKPV----- 2278
QY 1980 KOGOSNGVVOVOQKVLGIIPSSGTSTSOQFTSFOPRTATVTTRPNTSGSGGTTSNSQVI 2039
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APSKSGRDSGPS--- 2310
QY 2040 TGFOIRGMYVIRPLDQSTLGAIIITPVWVOPG---APQVMTQITIGQPVSTAVSA 2095
Db 2311 ---RPAQOPLSRPIQSPGRNS-----ISPRNGISPPNKLSQLPRTSSPSTA--- 2354
QY 2096 PNTVSTPPOGKSLTS-----ATSTSNIOSSASQPPPOGOQVKTMAQLTOLTQGH 2146
Db 2355 -STKSSSGSKMSTSPGRMSQONLTQGTLSKNASSIPSESS-----KGLNOMNGN 2408
QY 2147 GGNQGLTVIIOGGQQTGDLQ-----LIPQGVTVLPFGQOIQMAAMPNGTVQRFLE 2199
Db 2409 GANKVELISMSTSTKSSGSDSEREPVLVROSTFIEKAPSPILRLKLESAS-----FE 2463
QY 2200 PLATTATTATTTTSTTSTAAGTGEORQSLSP-----QMOVHOD-----KTLPAQSS 2248
Db 2464 SLSSSSPASPPTSQAOTPV-----LSFSLPDMISTHSSVQAGCKRLPPLSP 2513
QY 2249 SVGPAAQOPOTAPSAAPQOTOPQSPAPQEPVOTQPEVOTQTVSSHPSEAOPTHAQSS 2308
Db 2514 TIEYNDGRPAKRHDIAHSHESPRLPINSGTWKR--HSKSSSLPVPSTWRRTGSS 2570
QY 2309 KPOVAASQOSNVQOSPVAVQSPQOTRIRPSPOLS--PGQOSQVQTTTSQPIPIQPH 2367
Db 2571 SSIISSASESESEKAKSEDEKHVNSISGT--KQSKENQVSAKGTWRKIKENEFSP----- 2622
QY 2368 TSLQIPSOQOPQOSQNOTLSSGOT 2396
Db 2623 -----TNSTSQTVSSGAT 2635

RESULT 14
US-08-821-355A-7
; Sequence 7, Application US/08821355A
; Patent No. 5851775
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,355A
FILING DATE: 20-MAR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UT
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5851775e
US-08-821-355A-7

Query Match 2.4%; Score 362.5; DB 2; Length 2973;
Best Local Similarity 17.3%; Pred. No. 8,6e-13;
Matches 409; Conservative 327; Mismatches 86; Indels 777; Gaps 98;

QY 373 WEALCKILEEMREIEIRHNDITEDLTNKARGSKSFIAANEILIESIRAKGDIDNVK 432
Db 699 WDMCAVSMKLNLIHSHKIM-----AMGS-----AALNLNMANRPAKTKDA-NIM 743
QY 433 SPEET-----EKDN-ETENDSKAEKNREFEEDOSLEKOSDDKTPDDDEQKSEPT 485
Db 744 SPGSLPLSLHVKKQKALELDAQHLSETFDNIDNLSPKASHRSKORHKOSLTGDYVFD 803
QY 486 EYDGKGSVSAINGD-----NTT-----NATSEPTSPSGRSPVC--- 521
Db 804 NRHDNRSDNFTNGMTVLSPYLNTVLPSSSSSGSLDSRSRSEKDRSLERENGIGLGN 863
QY 522 --LSETPDSSNMAEKVASELPQ--DYVEEPN--KTCESSNTSATTTTSI-----QPNLEN 570
Db 864 HPATENGTSGSKGLQISTAAQIAKYMVEVSAIHTSQEDRSSGSGTTLHCVTDERALR 923
QY 571 SNSSSELNSSQESAKAADPENGERSHTPVSIQETVGDFTSEKSTGLSSPGAGK 630
Db 924 RSSAAHTHSNTYNTKTS---ENSNRFCSPFYA---KLEYKRSSNDLSNSVSSDGYGKR 976
QY 631 ASSGTRITRLRNPDKSLQSKS--QOVAAAHBANKL-FKEGKEYLVVN----- 677
Db 977 GQMKPSIESTSEDESEKFCGYQYPRADLAKIHSAHNDMDNDELDPPIVYSLKYSDEL 1036
QY 678 SGEISRLSTKKEYIMKGINNYFKLGQEGKYRYHNOYSTNSFALNKHQREDHDK--RR 736
Db 1037 NSGRQSPSONEMARPKHIIIEDEIKQSEQRQSR---NQSFT--YPVYTESTDKNLKFOP 1091
QY 737 HLAHKFLTPAGFEKMGVSHGSKVLITSLRLTLTLOLENNITSSFPHPMAHRAWIK 796
Db 1092 HFGQOECVSPYSRSGANGS-----ETNRVGS-----NHGINQNV 1125
QY 797 AYOMCSKPREFALALALECAVVPVVLPIWREFLHTRLHRTSIREKEKVKKKKK 856
Db 1126 SQSLQED-----DYEDKPTNISKRY 1147
QY 857 QEEETMQATWVKYTPPVAKHQAQKGEYRYVYGGWSMISTHYRFPRLPGNTVY 916
Db 1148 SEER-----QHEDDERPTNY-----SL 1164
QY 917 NYRKSLEGTKNNMDENDES-----DKRCKSRPKKIKIEPDEKDEV 959


```

Db 1165 KNEE-----KRAVDPIDISLKYADDIPSSQOKSEFSKSSGOSKTEHSSSSSTNTST 1220
QY 960 KSDAKAGADQEMDISKITEKKDQVELLSDS-----DKP-CKEEMPEV 1005
Db 1221 PSSNAKR---QNOHLHSSASOSRSGOQKATCQVSSINQETIOTYCEVDPTICFSCSSL 1277
QY 1006 DDDMKTESHVNCQESSQ-----VDVNVSGEFLRT-----SYKKRTSSKL 1047
Db 1278 SSLSAEDEIGCNOFTQOADSANTLQIAIEIKKIGTSAEDPVSEVPAVSQHPRTKSSRL 1337
QY 1048 DGLLEIRIQFLLEKORLEKIKLEGIGIGITSTNSSKN-----LSESPV-TRAKES 1101
Db 1338 QG-----SSLSSEARHKAVEFSSGAKSPSSGAQTPEKSPPEHYVQETPLMFSCSTV 1390
QY 1102 COSDSKROESPANNDQPEDILQCSQSDSVLRMSDPSHTNKLTPKRDVLDVIRS 1161
Db 1391 SSLDSFESHSIASSVQSEF-----CSGAVSGIISPDLDPSDGTMPFSR----- 1435
QY 1162 PETKCPKONSINDIEKYVDLASRQCEPTKSKTKGNDFIDSKLASAD----- 1211
Db 1436 SKTPPPPOTAQTKRE-----VPKNKAPTAEKRESG-----PKOAVNAAVQVQVLP 1483
QY 1212 DGTJL-----CKNK-----KPLIQESDPIV-----SSSKALHSSVK 1246
Db 1484 DADTLHFTESTPDGFSOSSLSALSLEDEPIQKVELRIMPPVOENDNGMETSEORP 1543
QY 1247 STNDRDAPLRAMDFEGKLGDSSENSTLENSDVTIODSSEEDMIQVNSNESISEQF 1306
Db 1544 ESENQ-----EKEAKTIDSEKD-----LDDSDDD----- 1570
QY 1307 RRREDVLEPLKCELVGSESTGNCEDRLPVKGTPEANGKPSQOKLEBPYKNCSDOI 1366
Db 1571 -----DIEFILE-EC-IISA-----MPTKSSR-KAKKPAQTASKLPVPARKPSOL 1612
QY 1367 -----KIKNTDKNNRESEKKGQRTSTFOINKDKMKPIYLKGBELKELSESRVSGNV 1423
Db 1613 PVYKLLPSONRLQPOKHVFTPG-----DDMPRY-----CV-----EOTPIINFST 1653
QY 1424 ERFVNNINKIIFENDIKSLTVKESALRPINGDVIMEDFNERNSSETKSHLSSDAEGN 1483
Db 1654 ASSLSTLIESPNEU-----AAGEVRGGAQSGEFKRTITTEGR--STDAQGG 1703
QY 1484 YNDSLETLPTKES-----DSTQTTTPSASCPS 1512
Db 1704 KTSV-TIPELDNKAEEGDILAEICINSAMPKSHKPPRVKIMQYQASASSAPNK 1762
QY 1513 NSVNOVEDMEITSEYKKTYS-----SPITSEESNLSNDF 1548
Db 1763 NOLDGKK--KKPTSPVKPIPONTERTVRVKNADSKNNLNAERVFSDNKDSKQNLKNS 1820
QY 1549 IDENGLPIKNENVNGE-----SKRKTIV 1571
Db 1821 KPNNDKLPNNEDRVGSPAFDSPHHTPIEGTPYPCSRNDSLSLDDDDVDVLSREK-- 1878
QY 1572 ITEVTMTSTVATESKTVIKVEKGQTVVSGSTENCAKSTVT-----TTTTTVK 1621
Db 1879 -AELRRAKENSEAKVTSHTELTSNQSANKTQAIKQPINRGQPKLLOKOSTPQGS 1937
QY 1622 LSTPSTGSGVDIISVKEGSKTVVTTVYDLSLTTGGTLVTSMTVSKVSTROKVKLMKS 1681
Db 1938 KDIPRGATD--EKLOMFAIENTPVCS--HNSLSLSLDIOENNNKENEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FVTKSTKSTIFVLBNDL-----K 1716
Db 1992 PPDOSGEPSKPAASGYAPRSFHEVDTPVCFSRNSSLSTISDSEDDULOECISSAMPKK 2051
QY 1717 KLARRGGIREVYFYNNAP-----ALDIMPYSPRPRTGIT-----WRYRL 1758
Db 2052 KPSRLKGDNE-----KHSPRNMGGIIGEDLTLDKDQRPDSEHGLSDSENFDMKAIQ 2105
QY 1759 QTVKSLAGVSLMLRLMLRMDMAAKVPPGGGSRFTEETSETETTTTITIRRDVGPYG 1818

```

```

Db 2106 EGANSTV-----SSLHQAAAAACL-----SRQASSDSD-----STLSLKS-----G 2141
QY 1819 IRREYCIIRKICIGVPEPTEKETPQORGLRSGSALRPKRPETPKOTGPVILITWAEER 1878
Db 2142 I-----SIGSPHILTPQOEKPFSTNKGPR--ILKPGKST-----LETKIIES 2184
QY 1879 -----LEIMEIRAEFAERVEKEKAQAEQAKRLEQOKP-----TVIATST 1919
Db 2185 SKGKGGKRYKKSLINGKVNSE-----ISQMKOPLQANPSPISRGRTMIHPG 2235
QY 1920 TSPSTSTTSPISPAQVMVAPISVTTGKMWLTITKVGSPAVTVTOQKNFNOTPATW 1979
Db 2236 VRNSSSTSPVSKKGPPLKTPASKSPSEQATTSBPAKPSV----- 2278
QY 1980 KOGSNSGVVQVQOKVLGILIPSSSTGTSOQFTSFQPRFATVTRPNTSGSGGTSNSQVI 2039
Db 2279 ---KSPSLVAPARTSQIG-----GSSK-----APRSRGSRODTPS----- 2310
QY 2040 TGPQIRPGMTVIRTPLOQSTLGAIRTRPYVWQPG---APQVMTQIIRGQPVSTA 2095
Db 2311 ---PPAQOPLSRPIQSPGRNS-----ISPGRNGISPPMKLSQLPRTSSPSTA--- 2354
QY 2096 PNTVSTPGOKSLTS-----ATSTNIQSSASQPPRPOGGQVKITMAQLTQLTQGH 2146
Db 2355 -STSSSGSGMSTSPGRQNSQNLTKQGLSKMASSIPRESAS-----KGLQNMNGN 2408
QY 2147 GGNGLTVVYQOGGQTTGOLQ-----LIPQGVTVLPBGQOOLMOAMPNGTVORFLFT 2199
Db 2409 GANKKVELMSSTKSSGSDSERPVLROSTFLKEAPSPILRRKLEBSA-----FE 2463
QY 2200 PLATTATTAATTTTSTTAAGTGEOROSKLSP-----QMOVHD-----KTLPPAQS 2248
Db 2464 SLSPSRPASPRTSOAQTEV-----LSPLPDMSLSTHSSVQAGGWKRLPPLNLSP 2513
QY 2249 SVGPAAKQQTQAPRARPOQTQPOSAPQEVQTOGEVQTOQTIVSSHVSEAPPTAQS 2308
Db 2514 TIEYNDGRPAKRHDIAHSHSESPSLPINSRGTWKRE---HSHSSSLPVPVSTWRRTGSS 2570
QY 2309 KPOVAASOPQSNVQOSQSPVRVQSPQOTRIRPSTPQLS--PGQOSOVQTTTSPQIRPIPH 2367
Db 2571 SSTLSASSSESEKAKSEDEKHVMSISGT--KOSKEQVSAKGVKRIKENEFSF----- 2622
QY 2368 TSLQIPSGOPOSOPOVOSSTOTLSSGQT 2396
Db 2623 -----TNSOTQVSSGAT 2635

RESULT 15
US-09-003-687A-7
; Sequence 7, Application US/09003687A
; Patent No. 5998600
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTION: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```



```

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/003,687A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/821,355
FILING DATE: 20-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A
REGISTRATION NUMBER: 32,145
REFERENCE/DOCKET NUMBER: 1107.05064
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 97430 BMB UT
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5998600e
US-09-003-687A-7

```

```

Query Match 2.4%; Score 362.5; DB 2: Length 2973;
Best Local Similarity 17.3%; Pred. No. 8,6e-13;
Matches 409; Conservative 327; Mismatches 856; Indels 777; Gaps 98;

```

```

QY 373 WEAELEKILEMEEEIHRHDDITEDLTNKARGSKSFLLAAANEELLESIFAKGIDIDNWK 432
DB 699 WDGAVSMKLNLIHSRHKM-----AMGS-----AALLNLNANRPAAKYKA-NTM 743
QY 433 SPEET-----EKDN-ETENDSKDAEKNEFEFDSLEKSDDKTPDDPEQKSEPT 485
DB 744 SPSSSLPLHVROKALEHLDQHLSETFDNIDNLSPKASHSKQRHKOSLYGDVYFDT 803
QY 486 EVGDKNYSANLGD-----NTT-----NATSETPSEGRSPVC-- 521
DB 804 NRHDDKRSDFNTGNMTVLSPLYNTVLPSSSSSSSGSLDSSRSEKRSLEEREGIGLGNV 863
QY 522 --LSETPSSNMAKKVASELPQ--DYPEPN--KTCESNTSATTTST-----QNLNEN 570
DB 864 HPTETEPGSSKRGLOISTIAOIAKAMEVSAIHTSOEDRSSGSTLTLHCVTDERNALR 923
QY 571 SNSSSELNQSOSAKAADDPENGEREHSHTPVSIQEIYODFTSEKSTGELSESPAGK 630
DB 924 RSSAAHTSHNTVFTKS---ENSNTCSMPYA---KLEYKRSSNDLSNVSSSSDGYGKR 976
QY 631 AGSGTILITRLNRPDSKLSQKS--QCVAAAHAEANKL-FKGGKVLVYN----- 677
DB 977 GQMKPSIESYSEDESKFCYGOYPADLAHKIHSANHMDNDGELDTPINYSILKYSDEOL 1036
QY 678 SOGEIIRLSTKREKVINNNYFKLGQKGYVYHNOYSTNFEALNKHQHPEDHKK-RR 736
DB 1037 NSGRQSPQONERAKPKHIIEDIKOSEORQR--NOSTT--YPYTTSTDDKHLKFGQ 1091
QY 737 HLAHKCCLTPAGEFKWNGSVHSGKVLITSLRLITOLENNIPSPSFFPHNMAASHRANWK 796
DB 1092 HFGQOCVSPYRSRGANGS-----ETNNRNGS-----NHGILQNV 1125
QY 797 AVQMGCKPEEFALALALECAVKPVVMLPIWREFLGHTRLHMTSIEREKEXVKKEKK 856
DB 1126 SOSLCQED-----DYEDDKPTNYSERY 1147
QY 857 OEEEFMQATWYKTFPVYKHQWVKQGEERYTGYGKSWISKTHVYFVFKRLPENTNV 916
DB 1148 SEBE-----OHEEERPTN-----S1 1164
QY 917 NYRKSLEGTAKNNMDENMDES-----DKRCSRSPKTKIKIPEDDSEKDEV 959
DB 1165 KYNEE-----KRHYDQPIDSLKAYDIPSSQKOSFSPKSSSGQSSKTEHMSSENTST 1220

```

```

QY 960 KGSDAKAGADONEMDISKITEKKQDYKELLSDS-----DKP--CKEEMPEV 1005
DB 1221 PSSNAKR-----QNDLHPSASRSRGQFOKATCAVSSINQETIOTYCVEDPTIFCSKSSL 1277
QY 1006 DDDMKTESHVNCQESSQ-----VDVNVSEGFHLRT-----SYKRTKRSKU 1047
DB 1278 SLSLSAEDETGQOTQOADSANTLQIAETIKETIGRSADPEVSEVPAVSQHPRTKSRLL 1337
QY 1048 DGLIERIKOFTLEEKQORLEKILBEGIKIGITSTNSSKN-----LSEPPV--TKAKKE 1101
DB 1338 QG-----SLSSESARHRAVEFSSGAKSPSKSGAQTPEPHYVQETPLMFSCSTVS 1390
QY 1102 QSDSMKROEQSPNANDOPEDLLQGCSDSPSVLRRASDSHTNKLKPYKDRVLDVDSIRS 1161
DB 1391 SLSLDSFESRSIASSVQEP-----CSGWSGILISPDLPDSGQMPPSR----- 1435
QY 1162 PETCKRKONSIENDIEKYSDLASRGOEPTKSTKCNDFIIDSKLASAD----- 1211
DB 1436 SKTPPPPOPTAQTKRE-----VPKNKAPTAEKREEG-----PKQAAVMAAVQRYVLP 1483
QY 1212 DIGTLI-----CKNK-----KPLQESDRTV-----SSKSALHSSVPK 1246
DB 1484 DADTLHPATESTPPDGFSCSSLSALSDEPFIQKQVLELRIMPVQENONGNETSEQPK 1543
QY 1247 STNDRDATPLSRAMDEEGKLGCDSESNSTLENSSDPTVSIQDSSSEDMIVQNSNESISEQF 1306
DB 1544 ESNENQ-----EKEAEKTIIDSKD-----LLDSDDD----- 1570
QY 1307 RTEDQVEVLEPLKCLVSGESTGNCEDRLPVKGTANGKKSQCKLEERYPNKCSDOI 1366
DB 1571 -----DIEILE-EC-IISA-----MPTKSSR-KAKKPAQTASKLPPPARKPSOL 1612
QY 1367 ---KLNTDKKNENRESKKGQRTSTFQINKDKPKIYLKGECLKEISERNVSGNV 1423
DB 1613 PYYKLLPSQNRLOPQKHVSTPG-----DMPRYV---CV-----EGTPINEST 1653
QY 1424 EPRVNNINKIIPENDIKSLTVKESAIRPFLNGVIMEDENERSSETKSHLLSSSDAEGN 1483
DB 1654 ATSLSDLTIESPPNEL-----AAGEVYRGAGQGEFEKRDITPTIGR--STDQAQCG 1703
QY 1484 YRDSEFLPSTKES-----DSTQTTSPASCPSS 1512
DB 1704 KISSV-TIPLEDNKAEBGDILAECLNSAMPKGSKPPRVKIMQVQASISSAPAK 1762
QY 1513 NSYQVEDMEIETSEVKKKYS-----SPITSEESNLSNDF 1548
DB 1763 NQDQKK--KKTPSPVKPIQNTETRYTRVKKNAKSNINNAERVFSDNNKSKQNNKNS 1820
QY 1549 IDENGULPINKENYNGE-----SKRTIV 1571
DB 1821 KQFNDKLPNNNEDVRGSAFDSPHHTPIEGTPYCFSRNDSLSLDFDDDDVDLSREK-- 1878
QY 1572 ITFVTTMTSTVATESKTVYKVEKGDKQTVASSENCAKSGVT-----TTTTVTK 1621
DB 1879 -AEIRAKAKENKESAEKVTSHTELTSMQSANKTQAIKAPINRGQPKPILOKOSTFPQSS 1937
QY 1622 LSTPTSGASVDIISVKEQSKTAVTTVTDSLTFTTGTLVTSMTVSKESYSTRDKVKLMKFS 1681
DB 1938 KOIDPRGAATD--EKLQNFATIENTPVCS--HNSLSLSLDIOGENNKKKEPIKETE 1991
QY 1682 RP-----KTRSGTALPSYRK-----FVTKSTKKSIFVLPNDL-----K 1716
DB 1992 PPSQSGEPKPOASGYAPKPSFVHEDTPVCFSRNSSLSTSIQSEDDLLQECISSAMPKKK 2051
QY 1717 KLAARKGIREVYFNNNAKP-----ALDIMPYSPRPPTGIT-----WRYVL 1758
DB 2052 KPSRLKGDN-----KHSPRNMGILGEDTLTDKLDIQRPSEHGLSDSSENFPMKALQ 2105
QY 1759 QTVKSLAGVSLMLRLMLASLRMDMAKAVPPGGGSHRTFSTETITTTTIIKRRDVPYG 1818
DB 2106 EGANSIV-----SSLHAAAAACL-----SKAASSSD-----SILSKS-----G 2141
QY 1819 IPEYCIARKIICPIGVETPKETPPQQRKGLRSSALRPKRPTPRQOTGPVILLETWAAEE 1878

```



```
Db 2142 I-----SLGSPHLTPQOEKPFISNKGPR--ILKPGEKST-----LETKKIESE 2184
QY 1879 J-----LELMEIRAPAEVEKEKAQAEQOAKKRLQOKP-----TVIATSP 1919
Db 2185 SKGKGGKVVYKSLITGKVRNSNE-----ISGQMKOPLQANMPISIRGRTMHIHG 2235
QY 1920 TSPSTSTTISPBAQVMVAPISGSVTTGTKWLTTKVGSPATVTPQONKNFHQTPATWY 1979
Db 2236 VRNSSSTSPVSKGPPPLKTPAKSPSEQOTATSPRGAKPSV----- 2278
QY 1980 KQGSNSGVVOQKVLGIIIPSTGTSQOTFSPQRTATVITRPMTSGSGTTSNQVI 2039
Db 2279 ---KSELSPVARQTSQIG-----GSSK-----APSRSGSRDSTPS--- 2310
QY 2040 TSPQIRPGMTVIRTPLOQSTLGAIRTPVMVPG-----APOQVMTQIRGQPVSTAVSA 2095
Db 2311 -----RPAGQPLSRPIQSGRNS-----ISPGRNGISPPMKLSQLPRTSSPSTA--- 2354
QY 2096 PNTVSTPQOKSLTS-----ATSTSNIOSSASQPPRPQOGQVKLTMQULTQLTQGH 2146
Db 2355 -STKSSGSGKMSYTSPPGRMSQONLTKQTLGSKNASSIPRESAS-----KGLNQMNNGN 2408
QY 2147 GGNQGLTVVIOGGQOTTGQLQ-----LIPQVTVLPBGQOLMGAAMPNGTVQRFLET 2199
Db 2409 GANKKVELRMSSTKSSGESDSRSERPVLVROSTFIKEAPSPTLRRKLEESAS-----FE 2463
QY 2200 PLATTATTAATTTTIVSTTAAGTGEOROSKLSR-----OMOVHOD-----KTLPPAOS 2248
Db 2464 SLSPSRPASPTRSQOQTFV-----LSPSLPDMSLSTHSSVQAGGWRKLPNLSR 2513
QY 2249 SVGPAKAQPOTAQPSARPOPOPOSPAPQPEVQTOPEVQOTTVSSHPSEAOPTHAQSS 2308
Db 2514 TIEYNDGRPAKRHDIAKSHSESPSRPLINRSGTWKRE---HSKHSSSLPRVSTWRRGSS 2570
QY 2309 KPQVAAQSQPQSVQGSQSVVRVQSPQTRIRPSPQLS-PGQSQVQTTSPPIPIQPH 2367
Db 2571 SSTLSASSSESEKASEDEKHVNSISGT--KQSKENQVSAKGTWRKIKENEFSP----- 2622
QY 2368 TSLQIPSOQPOPOPOVQSSQTLLSSGQT 2396
Db 2623 -----TNSTSQTVSSGAT 2635
```

Search completed: November 20, 2002, 16:32:15
Job time : 105.976 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:05 ; Search time 74.106 Seconds

(without alignments)
5227.100 Million cell updates/sec

Title: US-09-698-295-1

Perfect score: 14971

Sequence: 1 MVSEEEEDGDAEETDSE.....KIKGFKASRSHNNKIQSTAS 2907

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
A.Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the entry being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14971	100.0	2907	21	AAV57452
2	14170	94.6	2781	21	AAV57453
3	2695.5	18.0	557	22	ABB94078
4	1488.5	9.9	576	22	ABB68735
5	1385	9.3	328	21	AAV54364
6	982.5	6.6	238	20	AAV04323
7	981	6.6	237	22	AAU19446
8	789	5.3	149	23	AAU78845
9	781	5.2	246	22	AAU31944
10	661.5	4.4	997	22	ABB62887

11	621.5	4.2	2768	22	ABB68397	Drosophila melanog
12	613.5	4.1	5533	22	ABB65772	Drosophila melanog
13	613.5	4.1	5560	22	ABB71160	Drosophila melanog
14	567	3.8	6815	22	ABB66811	Drosophila melanog
15	562.5	3.8	3257	22	ABB67502	Drosophila melanog
16	561	3.7	3080	22	ABB64877	Drosophila melanog
17	551.5	3.7	2897	22	ABB58514	Drosophila melanog
18	503.5	3.4	4498	22	ABB58515	Drosophila melanog
19	491	3.3	2137	23	ABP39618	Staphylococcus epi
20	474	3.2	3726	22	ABB63947	Drosophila melanog
21	462	3.1	2586	22	ABB66878	Drosophila melanog
22	458.5	3.1	3266	21	AAV42491	Human OREF ORF2255
23	449.5	3.0	3111	22	ABB60327	Drosophila melanog
24	444	3.0	2977	22	ABB69480	Drosophila melanog
25	443	3.0	2344	22	AAU37120	Staphylococcus aur
26	441.5	2.9	3696	23	ABP40235	Staphylococcus epi
27	439.5	2.9	1878	22	AAV40239	Human polypeptide
28	439.5	2.9	3256	21	AAV50976	Human cell cycle p
29	435	2.9	1727	22	ABB59554	Human protein sequ
30	433	2.9	1795	22	ABB69806	Drosophila melanog
31	432	2.9	2665	22	ABB28314	Human peptide #965
32	432	2.9	2665	22	ABB33490	Peptide #996 encod
33	432	2.9	2665	22	ABB18950	Protein #949 encod
34	432	2.9	2665	22	AAV54270	Human brain expres
35	432	2.9	2665	22	AAV66665	Human bone marrow
36	432	2.9	2665	22	AAV14533	Peptide #987 encod
37	432	2.9	2665	22	AAV26950	Peptide #987 encod
38	432	2.9	2665	22	AAV20259	Peptide #941 encod
39	432	2.9	2665	23	ABG36319	Human peptide enco
40	429.5	2.9	3201	22	ABB62899	Drosophila melanog
41	427	2.9	2951	22	ABB60291	Drosophila melanog
42	423	2.8	2285	22	ABB63057	Drosophila melanog
43	417.5	2.8	1878	19	AAW81170	Human BAZ2-alpha p
44	417	2.8	2441	21	AAV18161	Plasmodium falcipa
45	414.5	2.8	3169	22	ABB64569	Drosophila melanog

ALIGNMENTS

RESULT 1
ID AAV57452 standard; Protein: 2907 AA.
XX AAV57452;
AC -
XX -
DT 22-FEB-2000 (first entry)
XX -
DE Human transcriptional regulatory factor SEQ ID NO:1.
XX -
KW Human; transcriptional regulatory factor; TCOAL; BLAST detection;
KW bromo-domain; cell proliferation; cancer.
XX -
OS Homo sapiens.
XX -
PN WO9957143-A1.
XX -
PD 11-NOV-1999.
XX -
PF 30-APR-1999; 99WO-DP02340.
XX -
PR 30-APR-1998; 98JP-0137631.
XX -
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX -
PI Jones MH;
XX -
DR WPI; 2000-052940/04.
XX -
PT N-PSDB; AA239032.
XX -
XX Transcriptional regulatory factor containing a bromo domain and gene
XX TCOAL encoding it

PS Claim 1; Page 48-68; 154pp; Japanese.

XX The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulatory mechanism. It binds
CC to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.

XX Sequence 2907 AA;

Query Match 100.0%; Score 14971; DB 21; Length 2907;
Best Local Similarity 100.0%; Pident. No. 0;
Matches 2907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSEEEEDDGAETDSEDEDEDEDDDDSDYPEMEDDDDDASYCRESSFRSHST 60
DB 1 MYSEEEEDDGAETDSEDEDEDEDDDDSDYPEMEDDDDDASYCRESSFRSHST 60
QY 61 YSSPPGRRKPRVHRPSPILIEKDIPLLEPKSSSEDLAVNEHIMNVAIYEVRNFGTV 120
DB 61 YSSPPGRRKPRVHRPSPILIEKDIPLLEPKSSSEDLAVNEHIMNVAIYEVRNFGTV 120
QY 121 LRLSPFRREDCAALVSOEECTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
DB 121 LRLSPFRREDCAALVSOEECTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNSTLY 180
QY 181 FIDGTMPEVLRVYCESDEKREHNVLPQOAEADYPYGPENKIKVLOFVDOFLTNARE 240
DB 181 FIDGTMPEVLRVYCESDEKREHNVLPQOAEADYPYGPENKIKVLOFVDOFLTNARE 240
QY 241 ELMSEGVIOYDHCRCVCHKLGLDLCETCSAVYHLECVKPRLEBVPEDMOCEYVAHKV 300
DB 241 ELMSEGVIOYDHCRCVCHKLGLDLCETCSAVYHLECVKPRLEBVPEDMOCEYVAHKV 300
QY 301 PGVYDCAVLAEOXNKPYIRHEPIGYDRSRKRWFLNRLLIIEEDTENEEKKIWTYSTKVQ 360
DB 301 PGVYDCAVLAEOXNKPYIRHEPIGYDRSRKRWFLNRLLIIEEDTENEEKKIWTYSTKVQ 360
QY 361 LAELIDCLDKYMAELCKILIEEMREELHRHMDITTEDLTNKRGSNKSFLAANEELIES 420
DB 361 LAELIDCLDKYMAELCKILIEEMREELHRHMDITTEDLTNKRGSNKSFLAANEELIES 420
QY 421 IRAKGDIDNVKSPETFEKDKNETENDSKOAEKNREFEEDOSLEKSDDKTPDDPEQOK 480
DB 421 IRAKGDIDNVKSPETFEKDKNETENDSKOAEKNREFEEDOSLEKSDDKTPDDPEQOK 480
QY 481 SEEPTEVGDKGNSVSNLGDNTNATSEETSPSGRSRPGCLSTPPSSNMAEKKVASSEL 540
DB 481 SEEPTEVGDKGNSVSNLGDNTNATSEETSPSGRSRPGCLSTPPSSNMAEKKVASSEL 540
QY 541 PODYEENPKTCESSNTSATTTSTIOPNLSENSSSSELNSSOSSEAKAADPENGERESH 600
DB 541 PODYEENPKTCESSNTSATTTSTIOPNLSENSSSSELNSSOSSEAKAADPENGERESH 600
QY 601 PVSIOEELVGDFTSEKSTGELSESPGAGKAGSSTRITTRLRNPDSLSQIKSOQVAAA 660
DB 601 PVSIOEELVGDFTSEKSTGELSESPGAGKAGSSTRITTRLRNPDSLSQIKSOQVAAA 660
QY 661 HEARKLEKEGEVLYVNSOGELISRLSTKKEVIMKGINNNYFKLGOEKKYVYVHYNOYSTS 720
DB 661 HEARKLEKEGEVLYVNSOGELISRLSTKKEVIMKGINNNYFKLGOEKKYVYVHYNOYSTS 720
QY 721 FALNRHGHREDHDKRRHLAKRFCLTPAGEFKWNSVHGSVLTISTLRLTITLOENNIPS 780
DB 721 FALNRHGHREDHDKRRHLAKRFCLTPAGEFKWNSVHGSVLTISTLRLTITLOENNIPS 780
QY 781 SFPHPNASIRANWIKAVOMCSKPREPALALALIECAVKPVVMLPIWREFLGHTRLHMT 840
DB 781 SFPHPNASIRANWIKAVOMCSKPREPALALALIECAVKPVVMLPIWREFLGHTRLHMT 840
QY 841 SIEREEKVKKKKKKKEEETMOQATWVKTFFPVKHOQVKOGGEYRVYGYGWSMISK 900

DB 841 SIEREEKVKKKKKKKEEETMOQATWVKTFFPVKHOQVKOGGEYRVYGYGWSMISK 900
QY 901 THVRFVYKPLPGNTNVNVRKSELECTKNMDENMDESCKRKSRSRPAKIKIEPDESEKDEV 960
DB 901 THVRFVYKPLPGNTNVNVRKSELECTKNMDENMDESCKRKSRSRPAKIKIEPDESEKDEV 960
QY 961 GSDAKGADQNMEDISKITEKKDQDVKEELDSDSDCKPCKEPMPVDDDMKTESHVNOES 1020
DB 961 GSDAKGADQNMEDISKITEKKDQDVKEELDSDSDCKPCKEPMPVDDDMKTESHVNOES 1020
QY 1021 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLLEIRIKOFTLEEKORLEKILEGGIKIGK 1080
DB 1021 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLLEIRIKOFTLEEKORLEKILEGGIKIGK 1080
QY 1081 TSTNSKNLSSEPVITAKKECCQSDSMRQSPNANNQDPEDLIQGCSDSSVLRMSDP 1140
DB 1081 TSTNSKNLSSEPVITAKKECCQSDSMRQSPNANNQDPEDLIQGCSDSSVLRMSDP 1140
QY 1141 SHTTNKLYPKDRVLDVSIKSPETKCPKONSIENTDIEEYSDLASRQEPYKSKTKGNDP 1200
DB 1141 SHTTNKLYPKDRVLDVSIKSPETKCPKONSIENTDIEEYSDLASRQEPYKSKTKGNDP 1200
QY 1201 FIDSKLASADDTLTLCKNKKPLIQEESDTYSSSKSALHSSVPKSTNDPAPLSRAM 1260
DB 1201 FIDSKLASADDTLTLCKNKKPLIQEESDTYSSSKSALHSSVPKSTNDPAPLSRAM 1260
QY 1261 DFEKGLGDSKSNSTLNSDPTVSIQDSSEDMIVONSNISISQFTRQDQVEVLEPLK 1320
DB 1261 DFEKGLGDSKSNSTLNSDPTVSIQDSSEDMIVONSNISISQFTRQDQVEVLEPLK 1320
QY 1321 CELVSGESTGNCEDRLPVKGTETANGKRPQOQKLEEBPVKCSQDQIKLMTTKKNNEIR 1380
DB 1321 CELVSGESTGNCEDRLPVKGTETANGKRPQOQKLEEBPVKCSQDQIKLMTTKKNNEIR 1380
QY 1381 ESEKKGRTSTFOJNGDKNPKIYLKGECLKEISESVYSGVNEBPKNYNNKIIIPENDIK 1440
DB 1381 ESEKKGRTSTFOJNGDKNPKIYLKGECLKEISESVYSGVNEBPKNYNNKIIIPENDIK 1440
QY 1441 SLVYKESAIRPFIINGVIMEDFNRNSESFTKSHLSSDAGEVNRDSEFLPSTKESDST 1500
DB 1441 SLVYKESAIRPFIINGVIMEDFNRNSESFTKSHLSSDAGEVNRDSEFLPSTKESDST 1500
QY 1501 QTTTPSASCEPSNVNOVEDMEIETSEVKYKVTSSPITSEESNLSDNDFIDENGILPKNE 1560
DB 1501 QTTTPSASCEPSNVNOVEDMEIETSEVKYKVTSSPITSEESNLSDNDFIDENGILPKNE 1560
QY 1561 NVNGESKRRTVITEVTTMTSTVATESTVYIKVEKGDKQIVVSSSTENCAKSTVTTTTVT 1620
DB 1561 NVNGESKRRTVITEVTTMTSTVATESTVYIKVEKGDKQIVVSSSTENCAKSTVTTTTVT 1620
QY 1621 KLSPTSGVSDIISVKEQSKTVVTVTDSLTGTLVTSMTVSKESTRQKVLKMF 1680
DB 1621 KLSPTSGVSDIISVKEQSKTVVTVTDSLTGTLVTSMTVSKESTRQKVLKMF 1680
QY 1681 SRPKTRSGTALDSYKRFVYKSTKKSIFVLPNDDUKLARKGSIREVYFNNYNAKPAALDI 1740
DB 1681 SRPKTRSGTALDSYKRFVYKSTKKSIFVLPNDDUKLARKGSIREVYFNNYNAKPAALDI 1740
QY 1741 WPYSPRPFGITWRVYLOVKSILAGVSLMLRLIMASIRMDMAAVPPEGSGSTRTEPSE 1800
DB 1741 WPYSPRPFGITWRVYLOVKSILAGVSLMLRLIMASIRMDMAAVPPEGSGSTRTEPSE 1800
QY 1801 TELITTEILIKRQVGPAGIIEFYCIKRIICPIGVETPKETPPORRKGJLSSALARKRPE 1860
DB 1801 TELITTEILIKRQVGPAGIIEFYCIKRIICPIGVETPKETPPORRKGJLSSALARKRPE 1860
QY 1861 TPQOTGVLIETWVAEDELMEIRAFERVEREKAQAVEQAKRLEQOKPTVIATSTT 1920
DB 1861 TPQOTGVLIETWVAEDELMEIRAFERVEREKAQAVEQAKRLEQOKPTVIATSTT 1920
QY 1921 SPTSSTSTISPAQKVVAVPISGSVTTGKRVLTITVGSATVTPQOKNRFQTEATWVK 1980

Db 1921 SPTSSTTSTISPAQKVMAPISGSVTTGTAKMLTTVGSFAVTFQONKNHOFATWAK 1980
Qy 1981 OGQSNQSVVVOVQKVLGIIIPSSSTGTSGQFTTSPQRTATVTIRPNTSGSGGTTNSQVIT 2040
Db 1981 OGQSNQSVVVOVQKVLGIIIPSSSTGTSGQFTTSPQRTATVTIRPNTSGSGGTTNSQVIT 2040
Qy 2041 GPOIRGKMTVIRTPLOOSTLGAIIITPVWVOPGARPOQVNTQIIRKOPVSTAVSAPNTVS 2100
Db 2041 GPOIRGKMTVIRTPLOOSTLGAIIITPVWVOPGARPOQVNTQIIRKOPVSTAVSAPNTVS 2100
Qy 2101 SPPGOKSLTSASTNSIQAASQPPRPQOGVQLTMAQLTQLGNGHNGGLTVVIOGOG 2160
Db 2101 SPPGOKSLTSASTNSIQAASQPPRPQOGVQLTMAQLTQLGNGHNGGLTVVIOGOG 2160
Qy 2161 QTTGQQLIPQGVTVLPGRGQQLMAQAMPNGTVQRFLEFPLATTATTAATTTTSTTAA 2220
Db 2161 QTTGQQLIPQGVTVLPGRGQQLMAQAMPNGTVQRFLEFPLATTATTAATTTTSTTAA 2220
Qy 2221 GTGEORQSKLSPQMOVHODKTLPPAOSSSVGPAAKAPQTAPSAARPOTOPQSPAQPEV 2280
Db 2221 GTGEORQSKLSPQMOVHODKTLPPAOSSSVGPAAKAPQTAPSAARPOTOPQSPAQPEV 2280
Qy 2281 QTPQEVQOTQTVSSHVSEAPQRTNAOSSKQVAAQSQPQSNVQSGSPVRYQSPQRTIRP 2340
Db 2281 QTPQEVQOTQTVSSHVSEAPQRTNAOSSKQVAAQSQPQSNVQSGSPVRYQSPQRTIRP 2340
Qy 2341 STPSQLSPQSQSQVQTTTSPQPIQPTTSLQIPSGQPOQSPQVOSTQTLSSGOTLNV 2400
Db 2341 STPSQLSPQSQSQVQTTTSPQPIQPTTSLQIPSGQPOQSPQVOSTQTLSSGOTLNV 2400
Qy 2401 SVSSPSPRPOQLQOPQPVQIAPVQOLQOQVQVLSQIQSQVVAQIQAOQSGVQOIKQLDLP 2460
Db 2401 SVSSPSPRPOQLQOPQPVQIAPVQOLQOQVQVLSQIQSQVVAQIQAOQSGVQOIKQLDLP 2460
Qy 2461 QIQOQSSAVQTHQIQNVVTVQAASVQOLQVQOQLRDQOQKKQOQLEINNTSPSKLITV 2520
Db 2461 QIQOQSSAVQTHQIQNVVTVQAASVQOLQVQOQLRDQOQKKQOQLEINNTSPSKLITV 2520
Qy 2521 EITQOVVKNHNAVIEHLKQKSMTPAEREENQMITVCNQVMKYLIDKIDKEKQAAKRR 2580
Db 2521 EITQOVVKNHNAVIEHLKQKSMTPAEREENQMITVCNQVMKYLIDKIDKEKQAAKRR 2580
Qy 2581 KRESEVQQRKSKONATKLSALLFKHKEQLRAELIKKRALDKDQIQIEVQBELKRDLIK 2640
Db 2581 KRESEVQQRKSKONATKLSALLFKHKEQLRAELIKKRALDKDQIQIEVQBELKRDLIK 2640
Qy 2641 EKQMLQQAQTAQAACRPVTPVLPAPRAPRPPRPVQVHTGLSTPTLPVAAQQRKR 2700
Db 2641 EKQMLQQAQTAQAACRPVTPVLPAPRAPRPPRPVQVHTGLSTPTLPVAAQQRKR 2700
Qy 2701 EEEKDSSSKSKKKMISTTSKETKIDTKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI 2760
Db 2701 EEEKDSSSKSKKKMISTTSKETKIDTKLYCICKTPYDESKFYIGCDRCQNMVHGRCVGI 2760
Qy 2761 LQSEAEILIDEYVCPQOSTEDAMTVLPLTEKDEGLKRVLSLQAHKMAWPLEBVDPN 2820
Db 2761 LQSEAEILIDEYVCPQOSTEDAMTVLPLTEKDEGLKRVLSLQAHKMAWPLEBVDPN 2820
Qy 2821 DADYDYGVIKPEPDLATMERRVQRYYEKLTEVAADMTKIFDNCRYNPSDSFYQCAEV 2880
Db 2821 DADYDYGVIKPEPDLATMERRVQRYYEKLTEVAADMTKIFDNCRYNPSDSFYQCAEV 2880
Qy 2881 LESFVQOKLKGFKASRSHNNKLOSTAS 2907
Db 2881 LESFVQOKLKGFKASRSHNNKLOSTAS 2907

RESULT 2
AAV57453
ID AAV57453 standard; Protein: 2781 AA.
AC XX
XX AAV57453;
XX

DT 22-FEB-2000 (first entry)
XX Human transcriptional regulatory factor SEQ ID NO:10.
XX Human transcriptional regulatory factor: TCOA1, BLAST detection:
KW bromo-domain; cell proliferation; cancer.
OS Homo sapiens.
PN MO9957143-A1.
XX 11-NOV-1999.
PD 30-APR-1999; 99WO-JP02340.
PF 30-APR-1998; 98JP-0137631.
PR (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA Jones MH;
PI WPI: 2000-052940/04.
DR N-PSDB; AA239033.
PT Transcriptional regulatory factor containing a bromo domain and gene
PT TCOA1 encoding it
PS Claim 1; Page 139-151; 154pp; Japanese.
XX The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulation mechanism. It binds
CC to hSNF2H, hSNF2L and NCOA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
XX
XX
SQ Sequence 2781 AA:
Query Match 94.6%; Score 14170; DB 21; Length 2781;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
Qy 1 MYSEEEEDDGAETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFSSHST 60
Db 1 MYSEEEEDDGAETQDSEDEDEDEDDDDSDYPEEMEDDDDDASYCTESSFSSHST 60
Qy 61 YSSTPGRKRPVHRPNSPILEEKDIPPLEPKSSDLMVNEHIMVYAIYEVLRNFTY 120
Db 61 YSSTPGRKRPVHRPNSPILEEKDIPPLEPKSSDLMVNEHIMVYAIYEVLRNFTY 120
Qy 121 LRLSPREFDFCAALVSOECSTLMAEMHVVLLKAVLREEDTSNTTGPADLKQSVSTLY 180
Db 121 LRLSPREFDFCAALVSOECSTLMAEMHVVLLKAVLREEDTSNTTGPADLKQSVSTLY 180
Qy 181 FIDGTMPEVLRVYCESDKEVHNHLLPYQEAEDYRQVPEVKIKVLOFLVDOFLTTNIARE 240
Db 181 FIDGTMPEVLRVYCESDKEVHNHLLPYQEAEDYRQVPEVKIKVLOFLVDOFLTTNIARE 240
Qy 241 ELMSEGVIOYDDHCRCVHKLGDLILCCETCSAVVYHLECVKPRLEVEPDEMQCEVCAHHV 300
Db 241 ELMSEGVIOYDDHCRCVHKLGDLILCCETCSAVVYHLECVKPRLEVEPDEMQCEVCAHHV 300
Qy 301 PGVTDCAVETQKMKRYTRHEPTIGYDSSRRKYWFLNRLLIEEDTENENKKTWYSTKVO 360
Db 301 PGVTDCAVETQKMKRYTRHEPTIGYDSSRRKYWFLNRLLIEEDTENENKKTWYSTKVO 360
Qy 361 LALIDLCDDYWEAEELCKILEEMREEIHRHMDITEDLTNKAQSKSFLAANEELIS 420
Db 361 LALIDLCDDYWEAEELCKILEEMREEIHRHMDITEDLTNKAQSKSFLAANEELIS 420
Qy 421 IRAKKDDIDNVKSPTEETKOKNETENDSKDAEKNREEPDQSLKQSDKTPDDPQOK 480
Db 421 IRAKKDDIDNVKSPTEETKOKNETENDSKDAEKNREEPDQSLKQSDKTPDDPQOK 480

Db 421 IRANKGIDNVKSPBETBKDNKNTENDSKDAEKNREPEFOSLEKSDDKTPDDPEOGK 480
 Qy 481 SEEPTEVGDKNSVANLGDNTNATSEETSPSGRSPVGLSETPDSNNAEKKVASEL 540
 Db 481 SE----- 482
 Qy 541 PODYPEENKTCESSTNTSATTTSIQPLNLENSSSSELSNSESASAKADDPENGERSHT 600
 Db 483 ----- 482
 Qy 601 PVSIOELVGDPTSEKSTGELSESPGAGKAGSGSTRITRLRNPDSKLSQKSOQVAAA 660
 Db 483 -----VGFPKSEKSGELSESPGAGKAGSGSTRITRLRNPDSKLSQKSOQVAAA 534
 Qy 661 HEANKLFEKEGEVLVYNSOGELSRSLSTKEEYIMKGINNYKLOEGERYHYHNOYSNS 720
 Db 535 HEANKLFEKEGEVLVYNSOGELSRSLSTKEEYIMKGINNYKLOEGERYHYHNOYSNS 594
 Qy 721 FALKKHQREDHDKRRHLAHFCLTPAGEFKWNGSVHSGKVLITSTRLTTOLENNIPS 780
 Db 595 FALKKHQREDHDKRRHLAHFCLTPAGEFKWNGSVHSGKVLITSTRLTTOLENNIPS 654
 Qy 781 SFPHPNASHRANMIKAYOMCSKPREFALALILECAVKPVVMLPIMREFLGHTRLRMT 840
 Db 655 SFLHPNASHRANMIKAYOMCSKPREFALALILECAVKPVVMLPIMREFLGHTRLRMT 714
 Qy 841 SIEREKEKVKKKKEKKEEFTMOATVVKYTFPVKHQVWKQKGEYVVTGSGMSWISK 900
 Db 715 SIEREKEKVKKKKEKKEEFTMOATVVKYTFPVKHQVWKQKGEYVVTGSGMSWISK 774
 Qy 901 THYVRFEVVKLGNTNNVTRKSLGCKNNMDEMDSKRCSSRPKTKIEPDSEKDEVK 960
 Db 775 THYVRFEVVKLGNTNNVTRKSLGCKNNMDEMDSKRCSSRPKTKIEPDSEKDEVK 834
 Qy 961 GSDAAKGADQEMDISKTTEKKDQVLELDSDPCKEPEMVEDDMKTESHYNCOES 1020
 Db 835 GSDAAKGADQEMDISKTTEKKDQVLELDSDPCKEPEMVEDDMKTESHYNCOES 894
 Qy 1021 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLERRIKOFTLEKORLEKIKLEGGIKIGK 1080
 Db 895 SQVDVNVSEGFHLRTSYKKKTKSSKLDGLERRIKOFTLEKORLEKIKLEGGIKIGK 954
 Qy 1081 TSTNSSKLSLSPYITRAKECCQSDSMARQEOSPNANNQDPEDLIOGCSQSSSVLRMSDP 1140
 Db 955 TSTNSSKLSLSPYITRAKECCQSDSMARQEOSPNANNQDPEDLIOGCSQSSSVLRMSDP 1014
 Qy 1141 SHTNNKLTPKRVLDDVSIRESPEKCPKONSIENTDIEEKVADLASRGOEPTKSKTGMDP 1200
 Db 1015 SHTNNKLTPKRVLDDVSIRESPEKCPKONSIENTDIEEKVADLASRGOEPTKSKTGMDP 1074
 Qy 1201 FIDDSKLASADDIGTLICKNKPLIOESDPIVSSSKSALHSSVPKSTNDRDAPPLSRAM 1260
 Db 1075 FIDDSKLASADDIGTLICKNKPLIOESDPIVSSSKSALHSSVPKSTNDRDAPPLSRAM 1134
 Qy 1261 DFEKRLGDSNSTLEKNSSTVSIQDSSSEDMIVONSNEISSEOFRIREDDEVLEPLK 1320
 Db 1135 DFEKRLGDSNSTLEKNSSTVSIQDSSSEDMIVONSNEISSEOFRIREDDEVLEPLK 1194
 Qy 1321 CELVSGESTGCEBRLPVKGTGEANGKKPSQOKLEERPYNKCSQIKLKNTTDDKNNENR 1380
 Db 1195 CELVSGESTGCEBRLPVKGTGEANGKKPSQOKLEERPYNKCSQIKLKNTTDDKNNENR 1254
 Qy 1381 ESEKKGORTSTFOJNGDKNKRILYKGECLKEISESRVSGNVEPKVNNIKIIPENDIK 1440
 Db 1255 ESEKKGORTSTFOJNGDKNKRILYKGECLKEISESRVSGNVEPKVNNIKIIPENDIK 1314
 Qy 1441 SLATKESAIRFINGDVIYMEFPNRSNSETKSHLSSDABGNTRDSLETPSTKESDST 1500
 Db 1315 SLATKESAIRFINGDVIYMEFPNRSNSETKSHLSSDABGNTRDSLETPSTKESDST 1374
 Qy 1501 OTTPPSACPPSNVNOVEDMEIETSEYKKTYSPTSEESNLSNDFIDNGLPINKNE 1560
 Db 1375 OTTPPSACPPSNVNOVEDMEIETSEYKKTYSPTSEESNLSNDFIDNGLPINKNE 1434

Qy 1561 NVNGESKRKVIETEVTTMTSTVATESKTVIKVEKGDKQTVVSSYENCAKSTVTTTTVT 1620
 Db 1435 NVNGESKRKVIETEVTTMTSTVATESKTVIKVEKGDKQTVVSSYENCAKSTVTTTTVT 1494
 Qy 1621 KLSPTSGGSVDIISVKEOSKTVVTTVTVDLSJTTGGTLVTSMTVSKSEYSTRDKVKLMKF 1660
 Db 1495 KLSPTSGGSVDIISVKEOSKTVVTTVTVDLSJTTGGTLVTSMTVSKSEYSTRDKVKLMKF 1554
 Qy 1681 SRPKKTSGTLALPSYKRFVNTSKSTFVLBPNDLKLARKGCIREVYFYNKAPLADI 1740
 Db 1555 SRPKKTSGTLALPSYKRFVNTSKSTFVLBPNDLKLARKGCIREVYFYNKAPLADI 1614
 Qy 1741 WYPSPPPTFGITWRNRLQOTKSLAGVSLMLRLMASLRMDMAKAPPGGSGSTRTESE 1800
 Db 1615 WYPSPPPTFGITWRNRLQOTKSLAGVSLMLRLMASLRMDMAKAPPGGSGSTRTESE 1674
 Qy 1801 TEITTTTELIRRDVGPYGIREFYCIKIKICIGVETPKETPTPQKGLSSALRPKE 1860
 Db 1675 TEITTTTELIRRDVGPYGIREFYCIKIKICIGVETPKETPTPQKGLSSALRPKE 1734
 Qy 1861 TPXOTGPVLIETWVAEELLELMEIRAPAEYREKKAQAVEQAKKRLBQKPTVIATSTT 1920
 Db 1735 TPXOTGPVLIETWVAEELLELMEIRAPAEYREKKAQAVEQAKKRLBQKPTVIATSTT 1794
 Qy 1921 SPTSSTSTISPAOKVWVAPISGSVTTGTKMVLTTKVGSPATVTFQONKNFHQTEATWK 1980
 Db 1795 SPTSSTSTISPAOKVWVAPISGSVTTGTKMVLTTKVGSPATVTFQONKNFHQTEATWK 1854
 Qy 1981 QGOSNSGVVOYQOKVVLGIIPSTGTSQOFTSFOPRATVYIRNTVSGSGTTSNSOYIT 2040
 Db 1855 QGOSNSGVVOYQOKVVLGIIPSTGTSQOFTSFOPRATVYIRNTVSGSGTTSNSOYIT 1914
 Qy 2041 GPOIRPGMTVTRTPLQOSTLGAIRTPVWYQAPACQVMQILRGOVPSTAVSAPMTVS 2100
 Db 1915 GPOIRPGMTVTRTPLQOSTLGAIRTPVWYQAPACQVMQILRGOVPSTAVSAPMTVS 1974
 Qy 2101 STPGOKSLTSAFSTSNISQSSASQPPRPOGGQVLTMAQLTQLTGHGNGGLVIVIOGG 2160
 Db 1975 STPGOKSLTSAFSTSNISQSSASQPPRPOGGQVLTMAQLTQLTGHGNGGLVIVIOGG 2034
 Qy 2161 OTTGQOLILPGVTVVLRPGQGLQMAAMPNGTVORFLTPLATATATASTTTTVSTTAA 2220
 Db 2035 OTTGQOLILPGVTVVLRPGQGLQMAAMPNGTVORFLTPLATATATASTTTTVSTTAA 2094
 Qy 2221 GTGEORQSKLSPQMVHODKTLPPAOSSSVGPAPAAQOPOTAQPSARPQOPQPSAPAEV 2280
 Db 2095 GTGEORQSKLSPQMVHODKTLPPAOSSSVGPAPAAQOPOTAQPSARPQOPQPSAPAEV 2154
 Qy 2281 QTOPREVOTOTTVSSHVDSEAOPTHAOSKPOVAASQOPQSNVGOQSPVRVQSPQTRIRP 2340
 Db 2155 QTOPREVOTOTTVSSHVDSEAOPTHAOSKPOVAASQOPQSNVGOQSPVRVQSPQTRIRP 2214
 Qy 2341 STPSQLSPGOQSOVOTTTSPRIPIQPHTSILOIPBQGPQSPQVQSSQTLSQCTLNOV 2400
 Db 2215 STPSQLSPGOQSOVOTTTSPRIPIQPHTSILOIPBQGPQSPQVQSSQTLSQCTLNOV 2274
 Qy 2401 SVSSPSRPOILOQOPQOVIAVAPOLQOOQVYLSQIQGQVAAQIOAQOSGVPQOIKILOPI 2460
 Db 2275 SVSSPSRPOILOQOPQOVIAVAPOLQOOQVYLSQIQGQVAAQIOAQOSGVPQOIKILOPI 2334
 Qy 2461 QIQOSSAVOTHQIOVNTVYVAASVOEQLORVOQLRDQOQKKQOQIEINVTBPSKLLIKV 2520
 Db 2335 QIQOSSAVOTHQIOVNTVYVAASVOEQLORVOQLRDQOQKKQOQIEIKREHNTLOASNOS 2394
 Qy 2521 EITOKOVYMKHNAVIEHLKQKSMTPABREBNCMIYCNOMKTYITLDKIRKEEQAAKKR 2580
 Db 2395 EITOKOVYMKHNAVIEHLKQKSMTPABREBNCMIYCNOMKTYITLDKIRKEEQAAKKR 2454
 Qy 2581 KREESVOKRSKONATKLSALLFKHKEOLRAEILKKRALLDKDQIOIWOBELRDLKIRK 2640
 Db 2455 KREESVOKRSKONATKLSALLFKHKEOLRAEILKKRALLDKDQIOIWOBELRDLKIRK 2514

XX		Drosophila melanogaster polypeptide SEQ ID NO 32997.
DE		
XX		Drosophila melanogaster polypeptide SEQ ID NO 32997.
KX		Drosophila; developmental biology; cell signalling; insecticide;
KM		pharmaceutical.
XX		
OS		Drosophila melanogaster.
PN		
XX		WO200171042-A2.
PD		
XX		27-SEP-2001.
PE		
XX		23-MAR-2001; 2001WO-US09231.
PR		
PR		23-MAR-2000; 2000US-191637P-
XX		11-JUL-2000; 2000US-0614150.
PA		(PEKE) PE CORP NY.
PI		
XX		Venter JC, Adams M, Li PWD, Myers EW;
DR		WPI: 2001-656860/75.
XX		N-PSDB; ABLI2838.
PT		
PT		New isolated nucleic acid detection reagent for detecting 1000 or more
XX		genes from Drosophila and for elucidating cell signalling and cell-cell
PS		interactions -
XX		
PS		Disclosure: SEQ ID NO 32997; 21pp + Sequence listing; English.
XX		
CC		The invention relates to an isolated nucleic acid detection reagent
CC		capable of detecting 1000 or more genes from Drosophila. The invention is
CC		useful in developmental biology and in elucidating cell signalling and
CC		cell-cell interactions in higher eukaryotes for the development of
CC		insecticides, therapeutics and pharmaceutical drugs. The invention
CC		discloses genomic DNA sequences (ABLI6176-ABLI3051), expressed DNA
CC		sequences (ABLI01840-ABLI6175) and the encoded proteins
CC		(ABBI7737-ABBI2072).
CC		The sequence data for this patent did not form part of the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences.
CC		
XX		
SEQ	Sequence	976 AA:
Query Match		9.9%; Score 1488.5; DB 22; Length 976;
Best Local Similarity		31.9%; Pred. No. 1.3e-67;
Matches	353; Conservative	139; Mismatches 295; Indels 319; Gaps 21
QY	4 EEEEDDGAEEFODSEDDDEMEDEDSDDPREMEDDDDDASVCTSSPFSSHSTYS 63	
	: : : : : : : : : : : : :	
Db	99 EEEHYKGSFGSSEKSKSNEDDMILTPSDDESLEVAENSESFVC---SFNONGV--- 152	
QY	64 TEGRRKPRVHBRSPILLE-KDIPLEPKSSSEDIMVPNEHINVAIEVLNFGTVLR 122	
	: : : : : : : : : : :	
Db	153 ---GRPPRPSPPEPVWLQGRQYAALDLPPSSDELFIANTHVLRALSIVEYLRRFRHWVR 209	
QY	123 LSPFFREDCCALLVSDEOCTIAEMKVIVLKAVLRERDSNTTFGGADLKDSVNSLTFL 182	
	: : : : : : : : :	
Db	210 LSPFFREDCCALLACEQSALLTEVHIMLLKLRLPERDQGTHGPLDKRDTVINISLYL 269	
QY	183 DGMTPEVILRVVCSCEDEK-----HNHLPYQAEDRYGVGVENKIKVLOEVLVDFTNTLA 238	
	: : : : : : : : : : :	
Db	270 DSITMPEVILRVSVESDXTEDRNVEHTLSITE--YYTGIDNNLEVLQFLSDQFLSNSI 326	
QY	238 REELNSEGYIQYDDHCRCVKHKLGLDCETGSANVHLECVKPPLEEVPEDEMOCLEVAVH 298	
	: : : : : : : : : : : :	
Db	327 RDMVMQEGRIHYDDHCRCVCHRLGDLCCETCAVVYHLECVDPMDNVPTEDMQCGLCRSR 386	
QY	299 KVPGLTDCAAEIQQKKRPYIRHEPIDGRSRRRYWFNLRLTIIEEDTENENEKKIWIYSTK 358	
	: : : : : : : : : : :	
Db	387 KTSGVYDVCLPEKGQVILRHDSLGVDRGRKIYFIARKIFLIEDQENT---CWITSIT 442	
QY	359 VOLAEILDLDKDYEAEALCKILEERREIRHRMTTEDTLTKAROSNKSFLAANAEEILI 418	

Db	443	SKILLISRLDAEELERLRLHSQITERDELEORMKLTETLTINHEKTKTSYV	494		
Qy	419	ESIRAKGCDIDNVKSPETERKDKNETENDSKAEKNREEFODSLEKDDTTPDDPEQ	478		
Db	495	-----ELBOEAKNLLLEKEVLEDD--EKDDAKSESQSIEG	528		
Qy	479	GKSEEPTEVGDKGVSANLGDNTNTNTSETSPSGRSPVGLSTTPDSSMMARKVYAS	538		
Db	529	TKKGE-----	533		
Qy	539	ELPDVPEEPKKTCESSNTSATTTSTIOPNDENSSSEELNSSQESAKAADPENGERS	598		
Db	534	-----	533		
Qy	599	HTPVSIOEIVGDTSEKSTIGELSESGAGKAGSGSTRITTLRLNPDKLSQYVA	658		
Db	534	-----EKKVTR-----EKNQTLT--	547		
Qy	659	AAHEANKLFEGKEVLVNVSGEISRLSTKEVIMKININNYFKLGOEKYRVYNNQST	718		
Db	548	-----NGLIHKLGMGEQFNYYNQST	570		
Qy	719	NSFALNKQHREDHDKRRHLAHKFCPLTPAGEFKWNGSVHGSKVLTSTLRLTTGLENNI	778		
Db	571	NPIALNKFORNEEDKRRHLSHKFSLTASDFKWMIGITMGTDNMITTLRQTLINFEESI	630		
Qy	779	PSSFPHWMAHRAHWIKAYOMCSKPEEFALAILCECAVPVYMLPIREFLGHTLIR	838		
Db	631	AASFALNTIMVYVKKIIMAAVAMNARBPSEFAVLLLFQASLSKSVFANVNHEDLGHITLOR	690		
Qy	839	MTSIEREKEKVKKKKKRQ--EEETMOQATVWKYTFPKHQVWKOKGEEERVYVGW	895		
Db	691	ITSAEREBKRLERKREKREDEEBRNRLAFNYIKYTLGLKHQWKOGEEYRVHGWGM	750		
Qy	896	SWISKT-----HYVRFPKLPNGTINWYKRSLEGTANNDENNDESDKRCRSKTKKI	950		
Db	751	LMLSSRRCGVARRAOPLTGHNRYHYTV--GEENDVNEIT-----LVDRITGRF	799		
Qy	951	EPDSEKDEKVSDAKAGADQ--NEMDISKTREKDDVKELLDSDSKPCKEPEAEVDD	1008		
Db	800	MOQCESSVNDQVCHYLPDQKKNVKEIDYTE-----	831		
Qy	1009	MKTESHVNCQESSOVDVYVNSSEGHNL--RTSYKKTKKSSKLDGLERRIKQFTLEKORL	1066		
Db	832	KINGH-----IDVSKALNAPGRITYSKVARKRSLDDLDRLRLKLAIVEOM--	877		
Qy	1067	EKIKLEGKIGIKGTSTNNSLNLS 1092			
Db	878	ASKIPSDMKLVVSSQNTNANSKQT 902			
RESULT 5					
AC	AAB54364	standard; Protein; 328 AA.			
XX					
DT	09-MAR-2001	(first entry)			
DE		Human pancreatic cancer antigen protein sequence SEQ ID NO:816.			
XX					
KW	Human: pancreas; pancreatic cancer; pancreatic cancer antigen;				
KW	detection; diagnosis; identification; cyrostatic; neuroprotective;				
KW	nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;				
KW	antiinflammatory; cardiact; gene therapy; chromosome mapping;				
KW	linkage analysis; tissue identification; tissue typing; forensic;				
KW	neural; immune system; muscular; reproductive; gastrointestinal;				
KW	pulmonary; cardiovascular; renal; proliferative.				
XX					
OS	Homo sapiens.				
XX					
PN	WO20005320-A1.				

XX 21-SEP-2000.
 PD 08-MAR-2000; 2000MO-US05989.
 XX PF 12-MAR-1999; 99US-0124270.
 XX PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX PI Rosen CA, Ruben SM;
 XX DR MPI: 2000-579444/54.
 XX DR N-PSDB; AAC99129.
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 PS Claim 11: Page 1272-1273; 1379pp; English.
 XX
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytosolic,
 CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 328 AA;
 Query Match 9.3%; Score 1385; DB 21; Length 328;
 Best local Similarity 66.4%; Pred. 5.6e-63;
 Matches 286; Conservative 0; Mismatches 17; Indels 128; Gaps 2;
 QY 406 NKSFLAANEELIESTIRAKKGDIDNVKSP--EETEKDKNETENDSKDAEKNEEFEDQSL 463
 DB 12 NVSLSGTNEELIESIRAKKGDIDNVKSPGTEETEKDKNETENDSKDAEKNEEFEDQSL 71
 QY 464 EKSDSDKTPDDDEPGKSEPTVEVGDKGNSVANLGDNTTNAISEETSPBEGSPVACLS 523
 DB 72 EKSDSDKTPDDDEPGKSE----- 90
 QY 524 ETPDSSNMAEKVASELPDVPPEPNKTCBSSNTSATTTISQPNLENSNSSSELNSQSE 583
 DB 91 ----- 90
 QY 584 SAKAADDPENGERESHTPVSIQEIWDFTSEKSTGELSPPGAGKAGSGSTRILTRLN 643
 DB 91 -----VGFSEKSNGLSSPPGAGKAGSGSTRILTRLN 125
 QY 644 PDSKLSQKSOQVAAAHAHANKLFKGEKYLIVNVSQGEISRLTKEVKIMKINNNTFKL 703
 DB 126 PDSKLSQKSOQVAAAHAHANKLFKGEKYLIVNVSQGEISRLTKEVKIMKINNNTFKL 185
 QY 704 GQSGKRYVYNQSTNSFALNKHQHRDHDHDKRRLAHKFCLTLPAGEFKMGVSHGSKVLFT 763
 DB 186 GQSGKRYVYNQSTNSFALNKHQHRDHDHDKRRLAHKFCLTLPAGEFKMGVSHGSKVLFT 245

QY 764 ISTLRITITOLENNIPSSFFHPMASHRANWIKAVQCKSPREFALALIECAVKPVVM 823
 DB 246 ISTLRITITOLENNIPSSFFHPMASHRANWIKAVQCKSKREFALALIECAVKPVVM 305
 QY 824 LPTWREFLGHT 834
 DB 306 LPTWREFLGHT 316

RESULT 6

AAY04323
 ID AAY04323 standard; Protein; 238 AA.
 XX
 AC AAY04323;
 XX
 DT 18-JUN-1999 (first entry)
 XX

DE Human secreted protein SEQ ID NO:69.
 XX

XX Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS.
 XX

OS Homo sapiens.
 XX

PN WO9910363-A1.
 XX

PD 04-MAR-1999.
 XX

PF 27-AUG-1998; 98MO-US17709.
 XX

PR 29-AUG-1997; 97US-0056271.
 XX

PR 28-AUG-1997; 97US-0056073.
 XX

PR 29-AUG-1997; 97US-0056247.
 XX

PR 29-AUG-1997; 97US-0056270.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
 XX

DR MPI: 1999-190585/16.
 XX

PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX

PS Disclosure; Page 18; 170pp; English.
 XX

XX AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
 CC AAX30145 to AAX30173 represent the secreted proteins encoded by the 29
 CC human genes. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions,
 CC e.g. by protein or gene therapy. Also pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new genes. Specific
 CC uses are described for each of the 29 genes, based on which tissues they
 CC are most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
 CC used in the exemplification of the present invention.
 XX

SQ Sequence 238 AA;
 Query Match 6.6%; Score 982.5; DB 20; Length 238;

Best Local Similarity 78.8%; Pred. No. 1.4e-42;
Matches 189; Conservative 7; Mismatches 19; Indels 25; Gaps 3;

QY 2689 PTLVASQKRRKEFEK-----DSSSKSKKKMISTTSKETK-----DT 2727
Db 3 PTPR-----KTPYDSKFKYIGCDLCTNMTHGCVGITEEAKKMDVYICNDCKRAOEGSSSE 58

QY 2728 KLYGICIKTPYDESKFYIGCDRCQNMWYHGRVGILOSEAELIDEVYCPQOSTEDAMVLT 2787
Db 59 ELVYICIRTPYDESOQFYIGCDRCQNMWYHGRVGILOSEAELIDEVYCPQOSTEDAMVLT 118

QY 2788 PLTEKDHGKLRVRLSLQAHMAMPLEPVDNDAPDYGVYIKPEMDLATMEERVQRRY 2847
Db 119 PLTEKDEYGLKRVLRSLQAHMAMPLEPVDNDAPDYGVYIKPEMDLATMEERVQRRY 178

QY 2848 EKLFEFVADMTKIFDNCRRYNSDPFYQCAEVLESFPYQKLKGFKASRSHNNKLOSTAS 2907
Db 179 EKLFEFVADMTKIFDNCRRYNSDPFYQCAEVLESFPYQKLKGFKASRSHNNKLOSTAS 238

RESULT 7
AAU19446
ID AAU19446 standard; Protein; 237 AA.

AC AAU19446;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human diagnostic and therapeutic polypeptide (DITHP) #32.
XX
KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder.
XX
OS Homo sapiens.
XX
PN W0200162927-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001MO-US06059.
XX

PR 24-FEB-2000; 2000US-0184693.
PR 24-FEB-2000; 2000US-0184697.
PR 24-FEB-2000; 2000US-0184698.
PR 24-FEB-2000; 2000US-0184768.
PR 24-FEB-2000; 2000US-0184769.
PR 24-FEB-2000; 2000US-0184770.
PR 24-FEB-2000; 2000US-0184771.
PR 24-FEB-2000; 2000US-0184772.
PR 24-FEB-2000; 2000US-0184773.
PR 24-FEB-2000; 2000US-0184774.
PR 24-FEB-2000; 2000US-0184776.
PR 24-FEB-2000; 2000US-0184777.
PR 24-FEB-2000; 2000US-0184797.
PR 24-FEB-2000; 2000US-0184813.
PR 24-FEB-2000; 2000US-0184837.
PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204325.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.

PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
XX
PA (INCY) INCYTE GENOMICS INC.
PI Panzer SR, Spito PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiter TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR WPI: 2001-502867/55.
DR N-PSDB: AAS31017.
XX
PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT
PS Claim 27; Page 418-419; 522pp; English.
XX
XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
CC therapeutic (DITHP) polypeptides of the invention.
XX
SQ Sequence 237 AA:

Query Match 6.6%; Score 981; DB:22; Length 237;
Best Local Similarity 87.5%; Pred. No. 1.6e-42;
Matches 182; Conservative 4; Mismatches 6; Indels 16; Gaps 1;

QY 2716 ISTTSKETK-----DTKLYGICIKTPYDESKFYIGCDRCQNMWYHGRVCG 2759
Db 30 VGITEEAKKMDVYICNDCKRAOEGSSSEELVYICIRTPYDESOQFYIGCDRCQNMWYHGRVCG 89

QY 2760 ILOSEAELIDEVYCPQOSTEDAMVLTPLTEKDYGLKRVLRSLQAHMAMPLEPVP 2819
Db 90 ILOSEAELIDEVYCPQOSTEDAMVLTPLTEKDYGLKRVLRSLQAHMAMPLEPVP 149

QY 2820 NDAPDYGVYIKPEMDLATMEERVQRRYKLEFVADMTKIFDNCRRYNSDPFYQCAE 2879
Db 150 NDAPDYGVYIKPEMDLATMEERVQRRYKLEFVADMTKIFDNCRRYNSDPFYQCAE 209

QY 2880 VLESFVQKLKGFKASRSHNNKLOSTAS 2907
Db 210 VLESFVQKLKGFKASRSHNNKLOSTAS 237

RESULT 8
AAU78845
ID AAU78845 standard; Protein; 149 AA.


```
XX AC AAU78845;
XX DT 18-JUN-2002 (first entry)
XX DE SCAN/KRAB protein SKAT-2 binding protein, 7A1.
XX KM SCAN/KRAB protein associated with a Th2 phenotype; SKAT-2; asthma;
XX KM allergy; hay fever; atopic dermatitis; allergic rhinitis;
XX KM Alzheimer's disease; neurological disorder; apoptosis;
XX KM T-helper response; vaccination; immune response; autoimmune disease;
XX KM infection; cancer; kruppel-like zinc finger protein; 7A1; human.
XX OS Homo sapiens.
XX PN GB2364051-A.
XX PD 16-JAN-2002.
XX PF 06-APR-2000; 2000GB-0008549.
XX PR 06-APR-2000; 2000GB-0008549.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Blanchard AD, Champion B, Page KR;
XX DR WPI; 2002-306794/35.
XX DR N-PSDB: ABK47549.
XX PT New SKAT-2 polypeptide transcription factor, useful for treating or
XX PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
XX PT acid, antibodies and modulators -
XX PS .Disclosure; Page 37-38; 47pp; English.
XX CC The invention describes an isolated SCAN/KRAB protein associated with a
XX CC Th2 phenotype (SKAT-2) polypeptide (I). (I), especially where expressed
XX CC in recombinant cells, is used to identify agents that modulate SKAT-2
XX CC activity, and may be useful therapeutically. The agents, also (I) and
XX CC the polynucleotides (II) that encode it, are useful for treatment and
XX CC prevention of diseases responsive to SKAT-2 modulation, specifically
XX CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
XX CC Alzheimer's disease (and other neurological disorders, particularly where
XX CC related to apoptosis); conditions involving T-helper responses, and
XX CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
XX CC or humoral responses, to reduce development of a Th1 phenotype, and to
XX CC manipulate the immune response in autoimmune diseases, infections and
XX CC cancer. (II) is also used for recombinant production of (I); as source
XX CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
XX CC detection of mutations and for monitoring SKAT-2 expression in
XX CC association with disease. This is the amino acid sequence of the peptide
XX CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
XX CC SKAT-2.
XX SQ
XX Sequence 149 AA;
SQ
Query Match 5.3%; Score 789; DB 23; Length 149;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 9
AAU31944
ID AAU31944 standard; Protein; 246 AA.
XX AC AAU31944;
XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #2435.
XX KM Human; vaccination; gene therapy; nutritional supplement;
XX KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US08656.
XX PR 18-APR-2000; 2000US-0552929.
XX PR 26-JAN-2001; 2001US-0770160.
XX PA (HYSE-) HYSBO INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20; Page 529-530; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising
XX CC the nucleic acids encoding the polypeptides and cells genetically
XX CC engineered to express them are also useful for producing the proteins.
XX CC The proteins are useful in genetic vaccination, testing and
XX CC therapy, and can be used as nutritional supplements. They may be used to
XX CC increase stem cell proliferation; to regulate haematopoiesis; and in
XX CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX CC immune suppression and/or stimulation; as anti-inflammatory agents; and
XX CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid
XX CC sequences of novel human secreted proteins of the invention.
XX SQ
XX Sequence 246 AA;
SQ
Query Match 5.2%; Score 781; DB 22; Length 246;
Best Local Similarity 67.8%; Pred. No. 2.9e-32;
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;
```


XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL12500.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at fcp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2768 AA:
SQ
Query Match 4.28; Score 621.5; DB 22; Length 2768;
Best Local Similarity 18.9%; Pred. No. 1.2e-22;
Matches 584; Conservative 448; Mismatches 1197; Indels 865; Gaps 133;
OY 78 PLEEKDIP---PLEPKSSEDLAVPNEHIMNVAIYEVLNREGTVLRSPFREDCA 134
DB 84 PIITCPVPADVPHSPETELSIPEK---GCSIEKFPYPSGAQVPSNKNKFCCLCYC 140
OY 135 LVSOEOTLMAEMHVLLKAVLREEDTSNTTFGPADLKDSVNTLYFIDGTMPEVLRV 194
DB 141 INNOTKCVMO-----ECGLH VDG-----CLPIY 163
OY 195 CESDKENHVLPIQEAADYRGVEKIKVLOFLVDOFLTNN----- 236
DB 164 NKGS-----CCPYRYSQDH-----ENELD-----FMDSTTTTTRPTGFLI 209
OY 237 -----IARELMSEGV-IQYDHCRCVCHL-GDLLC-----CE-----TCSAY 275
DB 210 PTTTDCIHGEITPADGSLGKNACBHCYCMRGDVCAYOCECEPMAANGKSCRAPAA 269
OY 276 ECVKPLPEVPEDEMOCVCAHKVPGVTDV-----AETQKNKPYIRHEPIGYSRR 329
DB 270 EGGCCPSNVYCEDSSSTTELVETTPESATSVPAKGHAELPKVDVLOEHIDDDKNKE 329
OY 330 KYFPLNRLIIEEDTENENKKTMYSTKVALLELDCLDKYV---EAECLKILEMRE 386
DB 330 TATIPSELSGSEIEEEEEKD---KATVAPVQVTD--EKDFSPEBSSTAGIPDSR- 383
OY 387 EHRHMDITEDLNKAGSKSFLAANEELISIRAKK---GDIDNVSPETEEDKN 442
DB 384 -----IDLPSSTEEKSESTEAEDIVKIVTPPEPGSGEEDVPKPSQIPEKEIT 434

OY 443 ETE-----NDSKDAEKNEE-----FEDQ-----SLEKSDOKTPDDDEPG 479
DB 435 EDELIKVTSAPAKASPEEEVYVATTSAPTEDEVKPTAGTISEEEEGKPTP---AEEG 491
OY 480 KSEF-----PTEVGDKNSVSNLGDVNTNATSEETSPSEGRS-----P 518
DB 492 SGEEKDVKYTAPEETEDEKPKTPSAVY---ASDEKDEQPKPSGSGDELDLKPPTAPT 548
OY 519 VGCSSETPDSSNMAEKKVASELPQDYPE-EPNKCESNTATTSIOPNLNENSSSEL 577
DB 549 AGATISASESEEDQDEBK-STPAPTSVDIDIEPAKPTESSEASG-----EGEDVAKET 599
OY 578 NSSQSESAKAADDPENGERESHPTVSI-QEEIV-GDFTSEKSTGELSPPAGKAGSGST 635
DB 600 TPAGEASIAEGEEIYKGTTPAGEPSSSEDEEIVKGTTPAESESESEDE----- 648
OY 636 RIITRLNRPDKLSQLSQVAAAHENKLFKRGKE--VLYVNSQGISLSTKKEVIM 693
DB 649 --LTKVTTTPAGEPSVAGEEELIETTPAGEPSIAGEEIVKVTTPAGE--SSIAGEEIVK 705
OY 694 KGNINNYFKLGOEGKYRVYHNOYSTNSFALNKHOREDDHRRRLAHFCLTPAGEEKM 753
DB 706 -----VTTPAGESSSGEEELIK-----VTTPAGESSSE 734
OY 754 GSVHGSKVLT-----ISTRLTITQLENNITPSPFHPNMAHRANWIKV-VOMCSKRE 806
DB 735 GDEEIVKSTPAGEPSISGEEDVIKATTSAPKSDIEGKKEPETATVEPAEEVEDPAK- 792
OY 807 FALALALECAVKNVVMPIRREFLGHTRLHRMTSIEEEKKAKKKEKOEETMOQA 866
DB 793 ---TTPIAEAEPEEPIAGPIPTDGIS-----GEEIVKGTTPQLEOPEISES 839
OY 867 TWVYTPPVKKQVKKQGEERYVGTGYGWSVTSKTHYRVFKPGMNNV-YKRSLEGT 925
DB 840 TEV-----PVAED-----DLSSSTSASAISSTEGV 865
OY 926 KNNMDNMDSDSKRCKSPKIKIEPDSEKD-----EYKSGDAKAGDON 971
DB 866 QDAASETTSSAPARAGKDEAATVPTAOKDDEVEDQATDLPEVDVQSTAKTTTTE 925
OY 972 EMDISKITEKKDQVKELSDSKPCKEPEVEVDOMK----- 1010
DB 926 QPKRESSFEADAEIEVTTSSPADKOEVEPAEPADKHKKDEEDVQATDLPKSDIGPPV 985
OY 1011 --TESHVNCOESSQVDVNVNSEGPHL-RTSYKKKTKSKLDGLLEBRIKOFTL-----E 1061
DB 986 VDTAATGQPTSDETATDKPVSYLEPVSQVESPSTAKVNRNDTEKETLPPSGEDQ 1045
OY 1062 EKORLEKIKLEGIGK-----IGKSTNSSKNLSSESVITKAREGQSDSMROQSPNA 1115
DB 1046 SSEPLPAMDLPAGIPRGEDCLVEGKTYANNITVATAPCVDSK--CISLVACQOMECK 1103
OY 1116 NNDOPE-----DLQCSQS---DSSVLRMSDPSHTNKLIPDRVLDVYSISPETKC 1166
DB 1104 LPEMLKCTVAAADLLDCCPTTIDESTESAKEKDEESTAK--PONKIDEDVEISLSTE-EI 1160
OY 1167 PKONSIENDIEBK-----VSDLASRQOEPTKSKTKGNDFFIDSKLASADDI 1213
DB 1161 PKDVIPTGITEQPLSHVKPPEELIQVTSYPAQDESTIAVDDKP--IDBS----- 1210
OY 1214 GTLICNNKPLIOESDPTIVSSKSAHLSVSKSTNDRDATPLSRANDFEKLGCDSESN 1273
DB 1211 ---AEDKKPIGESEED-----SKPIDESEEDKKPVE----- 1238
OY 1274 STLENSSDVYSIQDSSSEDMIV-----QNSNDSISBOFTTRBOD---VEVLEPLKCE 1322
DB 1239 ---ESAEDKKRVEDSESEKKEPLPTVIPASELEKSEKBEDEKKTADPAAPTEQPEATTTPA 1295
OY 1323 LVSGESTGNCEDRLPVAGTEANGK---KPSQOKKLEEPVAKGCSDOI---KLKVTYTDKN 1376
DB 1296 QIADTAKEVDKDLATTISAPVSGEDELKPADEKKRTE-----TAQIIPDAEITPASTDEPE 1349

QY 1377 NENR-----ESEKKQRTST--FOINGKDNKPKIYLKGECLKEISSESRVSGNVEKRVNNI 1430
 Db 1350 STEELPTVLDLKKPREDSTIKGTAEPSDKVPEVPTASTENIEESDKFTTVAAPKISAS 1409
 QY 1431 NKIIP--ENDIKSLTVE-----SAIRPFINGDVI-----MEDNERNSSSTKSL 1474
 Db 1410 DETEPEEDLVPAETPEIESEFEVSTKPPAVOGPPLPLAPAEKKPVDAETSTADI 1469
 QY 1475 LSSSDAEGNVRDLSLTLPSTKESDSTQTTTP--SASCPSNSNVOQEDMEIEFSEVKKVTLS 1533
 Db 1470 STEPSAEVEKEASGETSSDNEIDAGASTTVPVPSADEDKPISTKTYE--ADCKFTTV 1526
 QY 1534 SPIT--SEESN--LSNPFIDENGLPT-----NKNENNGESK-----RKT 1570
 Db 1527 APADDEESNLPKLPDIFEEA--PVAVTTAAPSKD--GEOKPVEVEEKPDEDQCKP 1582
 QY 1571 VITEVMTSTVATESEKIVKEGDKQTVVSTENCAKSTVTTTTVTVKLSTSTSGS 1630
 Db 1563 IEDETSTTS-----SENEIEPSRATTIAPSKEE--PSEPSTGAPTKDEPAEPTDAP 1635
 QY 1631 VDIISVKQOSTVVTVTVDLSLTGTGLVTSMTVSKKEYSTRDKVLMKFSRPKKTRSGT 1690
 Db 1636 ESDES--KETPESEVPTTVA--GEKIPTSITPPDEPTATSAVAKPDEDVEKETST 1690
 QY 1691 ALPSYRKFTVSTKKSIFVLNDDLKILARKGIREVYFVYNNKAPALDIMPSPRPTF 1750
 Db 1691 EIPDAPASSSEEDNS--STDQIPS-----EVP-----EKKP----- 1720
 QY 1751 GITWRKRLQTVKSLAGVSLMLRLMASLRMDMAKVPDGC--GSTRTSETETITTE 1807
 Db 1721 -----ETPAQTPBEDDIYGATAPRTSDVPPVQ 1749
 QY 1808 IIRKRDVGPYIREYCIRKILICIVPEPTKEPTPQOK--GLSSALIRPRPETPKQTG 1866
 Db 1750 RLPE-----EVLAEIPOPSTETGIRKQOEDTTAAPSIDRK-- 1783
 QY 1867 PVILETVAEELEIMELIRAFAEVREK--EKAQAVEQ-----QAKKRLQOOKPTVIAST 1919
 Db 1784 -----EPVYTELDEBATVAPRISEKDEKTEEEKPEVQKPGEEBSEEEKKPIEDQVST 1839
 QY 1920 TSPSTSTSTTISPAQKVMAPISGSVTTGKRMVLTITVGSFATVTFQONKNFHOTFAMV 1979
 Db 1840 EGPVSTASEAGSTES-----SEEVKPSTEGEVAEKPED----- 1873
 QY 1980 KQGGSNS--GVYVOQKVLGIIIPSTG---TSQOTFTSHQPTATVTR--PNTSGSGGTT 2034
 Db 1874 KQPSSTAQAPVETIPEISTELPADQDKPTESEAVDSEDETSAPSDEKIPSVSG----- 1927
 QY 2035 NSQVITGPQIRPGMTVIRTPLQOSTLGKATIRFPVMVOPGAPQOV--MQOIIROGPVSTAV 2093
 Db 1928 --EEVEBEV-----TTASFOAAEDELKTPAESEPSSTDKVPEPEYOKPEDETKAD 1977
 QY 2094 SAPNTVSTPGQKSLTSATST--SNIQSSAQPPRPOGQVKTMAOLTQLOQHGCG 2148
 Db 1978 ETPESVTQVSDVATSTSAVAGDIEKDEQATASP--EBEEETKPIIAAELIPQ----- 2031
 QY 2149 NQGITVIVIOQOGQTTGQLOLIPQGYVLP--GEGQOLMQ--AAMPNGTVQRFLEPIATTA 2205
 Db 2032 -----PSEKEPEVDE--QEVESGTRKATPAESDQPIDEIAPATSGIDE-----ASTA 2076
 QY 2206 TTASTTTTTVSTTAAGGEGOROSKLSPOQVHODK-----TLPPAQSSSVGAKAPQOT 2259
 Db 2077 APKREESTIVASAA-----SP--AVHDEIKDVTTTQVADKEKVAAPQDDTKT 2123
 QY 2260 A-----QPSAR--PQPTQOPQSPAPEVQOTQPEVOT---QTVVSHNPS-----EAQPTH- 2304
 Db 2124 SIDVSTDSPTAODEKQDEKAPVAPATTVSSPTADSAADSSTPVEKPSVEIDTKRMD 2183
 QY 2305 --AQSSKPVQA--AQSQPQSNVQGSQPVYV--QSPSQTRIRPS-----TPSQLSFGQOQSQ 2353
 Db 2184 IMSQITIAPTHADGAASTSTDEDDQAPVTVSPQAEKTPVSPAPQDSDKTPSSAPQDADE 2243
 QY 2354 VQTTTSG-----PIPIQPHTSIQIPSGQOPQSQPOVQSTQTL-----SSQOTLNQVS 2401

Db 2244 IPATATPLDDNKIPATVAVPQTDGVPATAPALPDEDEKIQTTAPALDEEKIPSTAAPLDDEK 2303
 QY 2402 VSSPSRQOLOIQOPQPOVYIANPQLOQOYVLSQIOSQVVAQIOAQOQSGVPOQIKLOLPQ 2461
 Db 2304 IPAPVSPVVFVEPSSSEKPAVSEYDGE-----ESTEPVHVHENTSTDEPTSDAKLKPTTS 2358
 QY 2462 IQOSSAVQTHQIQNVVTVQAA-----SVQEQLOQRVQOLRDQOQKKKQOQIEINVTPEKL 2516
 Db 2359 APATSESPATIEAIVPETAPAELEKEVPEKATQEPLEKETPEKATBQPELEKETPEKA 2418
 QY 2517 LIKVEI-----IQOVVYMKHNAVIEHLKOKSMTPAER-----EENQRMIV 2557
 Db 2419 TEQPELEKETPEKATQEPLEKEVTDKATQEPESVDEKTTPEPVKPSLDSTDEDE----- 2474
 QY 2558 CNQVAKYITLDIKDEKKAARKKRESEVEOKRSQONATKLSAL----- 2601
 Db 2475 -----ESVESEESADKDKKNETEDTDKNEBEVPAVAVSEIQQSEEAVPPTG 2525
 QY 2602 --LFKHKEQLAEILKRRALLDKLOIEVQELKRDLIKKEKDMLOLAQTAVAAPCP 2659
 Db 2526 HPLPPH---LASSITTPAVADR-----VGE-----DEENTTVKLSSTTTSTTESP 2570
 QY 2660 VT-----PVLPAAPAPPSPPPPGVQHTGLSTPTLPVYASQKRRKEEKDSSKSKKKM 2715
 Db 2571 VTSAPSTTTVASOQOQPIPPPPYG--HA-----PEYEDEYDEEVEYFGGTCRYAGKL 2620
 QY 2716 ISTSKETKDKTKLYCICKTKTYDESKFYIGDRCQNMWHGCVGILDS-----EHEL 2767
 Db 2621 YVSAQOILRRDDPCDFC-----FRSDITLOQSCPPPIAGCHEEP 2661
 QY 2768 IDEYVCPQO-----STEDAMTVLTP 2788
 Db 2662 ISGFCPCPRYECPVSAVLANITTTSTTTSTLTP 2695

RESULT 12
 ABB65772
 ID ABB65772 standard; Protein; 5533 AA.
 XX
 AC ABB65772;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 24108.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX
 PN MO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-0509231.
 XX
 PR 23-MAR-2000; 2000OUS-191637P.
 PR 11-JUL-2000; 2000OUS-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL09875.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
 XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB137137-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5533 AA:

Query Match 4.1%; Score 613.5; DB 22; Length 5533;
Best local similarity 19.5%; Pred. No. 8.2e-22;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
QY 305 DCVAETIQKPKYTRHEPIGYDRSRKYWFILNR-----RIIIEEDT 344
DB 1912 DTADKAEKNO--RHEKEKKEQEKREKDLKQVEREKREKKAQOEEREKEDKAKKEEK 1968
QY 345 ENENEKKIWTSTKYQALIELDLDYMAELCKTIEEMREELHRIMDTT-DLTAKAR 403
DB 1969 ERERREKKAQEDREK-----KEREERELRE--KEGRDKEQKEKEIREKDLREKEQ 2015
QY 404 GSNKSEFLAANEELIESIRAKKGDINVKSPETEKNTEENDSKDAEKNREEFEDQSL 463
DB 2016 -----RERONREKELRKDLREKEMRKEDOREKELHEKQOREHEHR 2057
QY 464 EKSDDKTPDDDE--QGSKEEPT-----VGDKGNSVANLGDNTNATSEETSPS 513
DB 2058 EKROSRRAMQVEQEGGRRELISYOKSKMDIAGEASLTALDCHKNENAMDTIAQGT 2117
QY 514 EGSPVPGCSETSDSSNMAEKKVASELPDQVPEEPNK---TCSSNNTSA-----TTSTI 564
DB 2118 PGASP-----STP--SDNTPFKERSKLSRNSPYRLHRRILSDEQNSAGCGSGGSSH 2170
QY 565 QPNLENSNSSSELNQSOSSESAAADDPENGERES--HTPVIOEETVGDFTSEKSTELS 622
DB 2171 QIHEDYVVRIRRENSQINISVHSSNQRLNDRSRKEKSSPFED-----KNSSSHIS 2223
QY 623 ESGGAKGASGSTRITRLRNPDSKLSQLKSQVAAAANANKLFEGKGVLYVNSOGET 682
DB 2224 RPHGCGSSASSSKHNHR-----RDKHNQKGSASISITNSI---EVVVDPISTQTK 2271
QY 683 SRLSTKEVYMKGNINNYFKLGOEGKYRVYHNQYSTNSFALN-----HQHR----- 729
DB 2272 HNINTSEELQSHO-----PAREKEREHFSSHANSSSRHRSKRDHHRREKRRHSA 2324
QY 730 -----EDHDKRRRLAHKFCILT---PAGFFKWNQSVHSGKVLITSLRLITOLENNIP 779
DB 2325 ESTNTDEEHTPOOHNRHRTISAAGSGSAGELSAATITSSGKLHGHHRHSVERKSSRG 2384
QY 780 SSGFFHFNMAHRANWIKAVOMCSKPREPALALALECAVAVVMLPIWREFLGHTRLHRM 839
DB 2385 DEGHSSSSSLRAKLMMLSSADSDDTDA-----SKKHSI 2419
QY 840 TSTERE-----EKEVYKKKEKKOEETETQOATVAVTYTFVKHQVKKQKEEYRVGYCG 894
DB 2420 FDLPPDCPNVSMYDKVAKRCKNMQOAEKKIKAKFS---OLQOSRAKKKRRSTSYDG 2474
QY 895 WS---WISTHYVRFPKLPGLNTNVVRKSLGTTKNMMDENMDESDKRSRSPKIKIE 951
DB 2475 DSDTEEREDQH-----RNSGSSSFHGR--YPLSSSDDDDEETIQRKRS----- 2517
QY 952 PDSKEDEVKGSAAKADONE-----MDISKITEKKDQVYKELLDSDSKPCKEEPNEVD 1006
DB 2518 -----SDSDAEHGQONOGASTLADANRVROMO-QNLRRLDGDGS---SED--EIR 2563
QY 1007 DDMKTESHVNCQSSQVDVYVNSEG-----FHLRTSY-----KKTKTSKLDGLL 1051
DB 2564 RNWAKSHFEGKRRNSSTRIASDESOSOPAPDLITIKOEHPIDAPQETIKREOLSDQEKFK 2623

QY 1052 ERIKQFTLEKORLEKIKLEGITIGKTSTNSSKNLSESPVITAKKEGQSDSMROEQ 1111
DB 2624 SRHDSNSISIER-----KLKTEREI-----KTLELGFVNSSEYTAGKLEKESPTERRKKKK 2675
QY 1112 SPNANNDQPEDLTIGGQSQSSS-----VLRMSDPS-----HTNKLYPKRVLDDVSI-- 1159
DB 2676 SKRR-----LKSSTADTSAAGPLVMTPLPSPIFDVHSSSECKTKDNFDNLTKEC 2727
QY 1160 -----RSPPTCKPKONSIENDIEEKVSDLASRGOEPTKTKGNDPEFIDSK 1206
DB 2728 SSIPLEISAGERRKK 2773
QY 1207 LASADDIGTLICRKNKPLIOESDPTIVSSKSLHSSVPKSTNDROATPLSR----- 1258
DB 2774 KLSGEENHRLKSKSKSKSMNSCMTKLYNSGA-H---PSTPSLAPTPPSAOSTAQTSK 2829
QY 1259 ---AMDPEKLGCDSESNSTLSENSDPTVSIQDSSEDMIVQNSNESISQDFTREDOVE 1314
DB 2830 RGEDKMEFIGIISDEEESQPFQEAETNKDILPSS-----VSTTGPIVSALQTKQ--- 2881
QY 1315 VLEPFLKCELVSGSTGCEDRLPYKGTGEANGK---KPSQOKLEBPVKKCSQDIKLN 1370
DB 2882 --EP-----STPVS-----KNEAHITQLTVEPEQOQDLER---SRLSGSSSS 2921
QY 1371 TTDKNNENRESEK---KQORTSTFQINGKDNKPKIYLKGECLKEISESRVVSNGVEPK 1426
DB 2922 HADREHRRRRKREKREKREKQOREQONQIHKSSK-----YETK 2959
QY 1427 VNNINKIIPENDIKSLTVKESAIRPPIINGQVIMEDFNERNSETKSHLLSSDAEGNRYD 1486
DB 2960 VDDDNSVDMDEAGRALAQ-----LMSDFPTKPISE----- 2990
QY 1487 SLETLIPSTK---ESDSQTTPASCPESNSVNOV-----EDMEIETSEKVKVTSPTIT 1537
DB 2991 --EATPSTAATYNSDMTDVRFPSDN--EDNNSVDMTKQGVASQEOQKSKDKKKKKR 3047
QY 1538 SEESNLSNDFIDENGLPIKNKENVNGESRKRTVITEVTMTSTVATESKTVIKVEGDK 1597
DB 3048 EEKQEKLL-----QQORRESLPVASTSSAPRPGKLTGVAVQAASK 3088
QY 1598 QT-----VYSTENCAKSTV-----TTTTTVTKLSPRSGSVDIISVKEQS 1640
DB 3089 HADQLDRAKHISPPYCKPSPSLPCLIGDDDDALHTPKAKPTTPSSRGNDGLTPSRKRP 3148
QY 1641 KTV-----VTTVTYDLSLT-----TTGGGLVTS----- 1662
DB 3149 RLISPIPKPTLIANSSTLSTQSAETPVSSGTVISSSALATTPPISSTAAGVSAAPGLDNP 3208
QY 1663 -----MTVSEKYST-----RDKVYLMKFSRPRK 1685
DB 3209 TSASAQCKKKESFIPGHQGLDRIRISQSAVQSIISAEFNSISLDNIDEKRIPIASPPRA 3268
QY 1686 TRSGTALPSTRKPVYVYSTK-----ST 1707
DB 3269 TKPLDKLEESKSRVTTISOETESAVSALGSEFQSTSTTSDLDGDEMSSVNELEPTL 3328
QY 1708 FVLPNDDLKTLARK-----GIREVVPYFNNAKALDIMPVSP-----RPTGKIT 1753
DB 3329 VIAEPDEALAKALETAGEPASILEP---EMEPERAEPPDDEALIESEPVVEVL 3384
QY 1754 WRYRL-QTVKSLAGVSLMLLMAASLRWMDMAKVPVGGGSTRFETETITTEILIKRR 1812
DB 3385 DPRELNAVQSLKHEDWM-----DIKADTPQSEBDQIDLT-DTEENDE---AD 3429
QY 1813 DVGPGYIGRFEYCIRKIIICPIGVPEPKETPTPOAKGLRSSALRRKRPETPKQGVYIET 1872
DB 3430 SSGP-----SKIDETVQSSSPKESISNNS-----PTPRETANIDI-- 3466
QY 1873 WVAEELELWEIRAFARVRYEKEKAQAVEQAQAKKRLQOKP-----TVIATSTTSPTST 1926
DB 3467 -----PNVESQPKLSMESTPQPSVITKLPLFLDTPKTVAGLPPSPVKIE 3510


```

QY 1927 TSTISPAQKVVAVPISGVS-----TTGTRKMLTTKVGSPAVTFQONKRFHQTATWVKOG 1982
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3511 PRTISKIQQPLVQVGVVTLAPRHSTGSGISANSVINIDLNVISSCNTASASTASAS 3570
-QY 1983 QSNNGVVOVQOKVLGITPSSTGTSQGTFTSFQF-RTATVITRPMTSSGGTTSN---SQV 2038
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3571 SISFSGPTASQNAW---FOASTPRKQGITPQOALRTQSLIMQPTTISIPEQTPIHFAPOM 3627
-QY 2039 ITGFOI-----RPG--MTVIRPLQOSTL---GKALIRT-----PYMVRGAPQ 2078
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3628 VLSQSHHPQDPGTVMGIRAPRSPHSLHSPGKGVAAQSRVLGQLSPVGRPWVQSP--SPQ 3666
-QY 2079 VMDQILNGQVSTAVSAPNTVYSTPGOKSLTSATSTNIOSSA---SQPRPQOGQV 2135
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3687 ---QVOOTQOQHALLITPQSSNISPLASPTTRVLSSNSPTTSVNVSYQPRNQ----- 3737
-QY 2136 MAQLTOLTOGHGNOGLTVVYIQGGQTTGOLILPGQTVLPRGGQOLMOAMNGTVQ 2195
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3738 -----VPRQPSRKSVAEQVOTPRQIMTI---PLQKMTPIQVPHNPT-- 3774
-QY 2196 FLTPPLATTAATTTTSTTAAAGTGEOQSKLSPOMQYHODKTLPRQSSSVBPAKA 2255
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3775 -LISKVTVTQVQATQSQVASSPRLGS-----LPRHKNVH---LMAQON-----Q 3816
-QY 2256 QPQT-AQPSARROPOTQPSAPQREVOQTOREVQOTTVSSHVSEADPTNAQSSKPOVAA 2314
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3817 QPVYAKMTAHNOHQNHQMOQFHMQMIQROQHMQ-----QQQLHQSQS--QITS 3862
-QY 2315 QSQPQSVQSQSVKQVQSPQSTRTKRPSTPQLSPQ--QSQVQTTTSPQRTPIQPTSLQIP 2373
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3863 APQNMQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQHQ 3918
-QY 2374 SQGQPSQSPQVQS---STQTLSSGQTLNOVSVSSPRQDQIQDQOP--QVLAQVQLQDQ 2428
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3919 -QHGVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3973
-QY 2429 -----VOVLSQISQVVAQIQAOQSGVPRQIKQLDPIQIOQSSAVQTHQIQNV 2476
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3974 PQGVGHLGSTSTIFASQOHNQSLPAR-----GVPRQ--QHPQQLSHSSPCKPNTLVSV 4024
-QY 2477 -VTVQASVQEQDQORVQOQLHQOQKQKQO-----QEIINV--NTP 2513
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4025 NQGVQPPALILRVGSHSQPQOQOQLPHQOQSSGHPHQOKQLSSPGANLPLQDRLVQDNTPR 4084
-QY 2514 SKLILKEIIL-OKQV---VMKHNVAIENLKQKSMTPAREENQRMIVQNMVKYITDKI 2569
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4085 -KIIVQHQIYAQNOVPRPQOGNAIHPQNGKQSTPRGH----- 4123
-QY 2570 DKEKQAKKRRKRESEVQKRSKONATKLSAL-----LPHKEQQLRAETLKRALIDK 2622
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4124 -VEFTPRMSAQKTSSESVVIRPTPTGLAVISANFVGSILTEENLIKISQPKODELIEQ 4182
-QY 2623 DLOEYQ-----BELKRDJIKK-----EKDMLQAOATAVAAQPPRYTVLPAPRAP 2670
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4183 DSK-EVUSDVWSAKEVINDVYIKLDTPLASKDKAKRAVEQAL-----AP 4226
-QY 2671 PPSPPPRGVQHTGLSTPLPVAQSKRKREKSSSKSKKKM 2715
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 4227 APIPNPQGNQ--SMAQETALPTTSMVSNNSNDHDEDETETRL 4269

```

```

XX OS Drosophila melanogaster.
XX PN NO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PK 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW,
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL15263.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 40272, 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABH5737-ABH72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 5560 AA:
Query Match 4.1%; Score 613.5; DB 22; Length 5560;
Best Local Similarity 19.5%; Pred. No. 8.2e-22;
Matches 546; Conservative 405; Mismatches 1013; Indels 841; Gaps 118;
-QY 305 DCVAEIQKRPYIHEDIGYDRSRKRYWFLNR-----RLIEEDT 344
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1912 DTADKARKNQ---RHEKKEKQREKDKLQYEREREKDKAQEREREKDKAKKEEK 1968
-QY 345 ENENEKIIWYSTVQVLAELIDCDQDYWEELCKILIEKREELHNMIDTE-DLTAKAR 403
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1969 EREREKKAQEDREK-----KERERELE--KEQRDKQKEREIREKDLREKEQ 2015
-QY 404 GSNKSFLAANEELIESIRAKKGDIDVKSPEETKNETENSKDAEKREFEQSL 463
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2016 -----RENDNREKELRDKDLREKEREKEREKELHNRKODREKRN 2057
-QY 464 EKUSDKTPDDPE--QKSEEPTE-----VGDKNSVSNLGNNTVANSSETPS 513
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2058 EKEQSRAMDEQEGRGGRMEELSSYQSKMDIDGFASSLTAIDCQHNKEMAMDTIAQGT 2117
-QY 514 EGRSPVGCLESTPDSMAEKKVASELPQVYREPKN---TCSSNNTSA-----TTTSI 564
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2118 PGASP-----STP--SDNTPERSRKLSRNSPVRLHRRLLSSQESNHSAGGSGGSSH 2170
-QY 565 QPNLENSNSSELNQSQESAKADDPENGRES--HTPVSIQEIYGDFTSEKSTGELS 622
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2171 QIHEDYVKRIRMEQNSQNIYSHNSQRLNDRDREKKEKSSFKED-----KNSSSHIS 2223
-QY 623 ESPGAGGAGSGSTRITRLRNPDKLSQLSQVAAAHANKLFKGEKVLVNSQGEI 682
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2224 RPHCGCGSSASSKNNHR-----RDKNHQKGSASSIETNSIT---EVVVDPIQTK 2271
-QY 683 SRLSTKEVIMKGININNYFKLQEGKRVYVNIQYSTSPALNK-----HQHR----- 729
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
 ABB71160
 ID ABB71160 standard; Protein; 5560 AA.
 AC ABB71160;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 40272.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

Dd	2272	HNLTSEEBLQSHQ-----PRKEERERHFSSHANSSSRKSRKRDHNNHREKKRSHVA	2324
Oy	730	-----EDHCKRRHLAKFCFLT---PAGEFKNGSVHGSKYLITSLRLTITTOLENNIP	779
Dd	2325	ESTYTDDEHPHOONHPRHRRISAAGSGABELSSAATNTSSGKLHHQHRRSERRKSSRGS	2384
Oy	780	SSFPHPMAHGRANWIKAYOMCSKPREPFLALALILECANRPMVLMPIREFLGHTRLHRM	839
Dd	2385	DEGHSSSKSLRAKLMMLSSADSDPTDDA-----SKRHSI	2419
Oy	840	TSIERE-----EKEKYKKKKKKOEBEETMOQATWVAYTFPVKHQVWKOGEEYRVATYGG	894
Dd	2420	FDIIDDOPNYSMTDKVAKRCKMNOQABEKKIKAFS-----QLOGRAKKRSTSDG	2474
Oy	885	WS---WISKTHYVRYPKLPGLNTNVYKRSLEGTKNMNMENDESDBKRCSPAKIKIE	951
Dd	2475	DSDFEFDROH-----RNGSSSFHGR--YGLSSDDDDDEEHORRIS-----	2517
Oy	952	PDSKDEVKSDAKGADQNE-----MDISKITEKKDOVKRLDSDSDCKCKEPMYD	1006
Dd	2518	-----SSDSDEHGGODMOGASTLADANRVRQMO-QULRRLCGDGS--SED--EIR	2563
Oy	1007	DDMKTESHWCOESSQVDVNVWSEG-----FHLRTSY-----KKTKSSKLDDL	1051
Dd	2564	RNVKXHHFKPKRNSNSTRIASDESSQAPRLTIQEHPIRPAOEIKRQLSDEBCKFK	2623
Oy	1052	ERRIKOTFLEBKORLEKIGEGIKIGTSTNSSKNLSSEPVITKAKEGCGOSDMROEQ	1111
Dd	2624	SRHDSNSIEBR---KLKTEREI---KTELADFYNSSEYTYGTLKEYSPETKRRHKK	2675
Oy	1112	SPNANNOQPEDLIQCGOSNS-----VLNMSDPB---HTNKLRYKRDVLDVSL--	1159
Dd	2676	SKRR-----LKSSSTAOTSAQPLVMTPLTPEFVHSSSBCKTKFDNFDLKTEC	2727
Oy	1160	-----RSPETKPKONSIENDIEEKVSDLASRGOEPRKSTKGNDDEFIDSK	1206
Dd	2728	SSIPLEISAGBRHKRHKRKKRKKRKLNMTEATVPR-----SPTTNDT-----SSE	2773
Oy	1207	LASADIDIGTLCKNKPRLIOESDITVSSKSAHSSVPKSTNDRDATPLSR-----	1258
Dd	2774	KLSEEHHRRLKSKSKSMONSCNTKIYNSGA-H---PSTSPSLPAIPTSPASPAQTSK	2829
Oy	1259	---AMOFEGKLOGDPSNSTLENSSDIYQSIDSSEBMDIYQNSMESISBOQPRREQOVE	1314
Dd	2830	RGEDKMEIFGILISDEESQFPOAETNKDILPSS-----VSTQPIVSAALQTYKQ--	2881
Oy	1315	VLEPIKCELVSGESTGCEDRLPVKSTENGK---KPSQOKLEERPKNCSQDIKLN	1370
Dd	2882	---EP-----STPRS-----KNEEHLQITVHEPQOQOLR---SRLSGSSSSS	2921
Oy	1371	TTDKNNENRESEK---KGQRTSFQINGDKNDPKRIYLKGBCLKEISESRVSGNEBPK	1426
Dd	2922	HADREHRRREKREKRREKRSQREOQNOHQSKS-----YETK	2959
Oy	1427	VNNINKIIPENDIKSLTVKXSALRPPIINDVIMEDFENENSETSHLLSSDAEGNTRD	1486
Dd	2960	VDDNSVDMDEAGALEAQ-----LMSFDTKRIS-----	2990
Oy	1487	SLEPLPSTK---ESDSQTOTTPASCPESNSYOV-----EDMELENTSEVKKYTSPT	1537
Dd	2991	---EXTPSTAAITYSDMTDVFRESDN--EDNNSVDMTKQGYKSQOQOHKSKOKKKKKRKS	3047
Oy	1538	SEESNSLNDPIDENGLPIKNKNVNGESKRKYTIVETVMTSPVATESKTVIKREKDK	1597
Dd	3048	EKEGEXKL-----QOORRESLPNVAISTSPAPPTGCKLTVNVQASK	3088
Oy	1598	QT-----VVSSTENCASKTV-----TTTTTVTKLISTSGSVDIISVKEQS	1640
Dd	3089	HADLOLDAKHISSPPYCKPSPSLPICGDDDDALHTPRAKPTPSSGNDGLPSPREKP	3148
Oy	1641	KTV-----VTTTVVDSLT-----TTGGLVATS-----	1662
Dd	3149	RLISPIRKTPLIANSTLSQSAETPVSQGVYISSSALATTPPTSSTAAGVSAAPGLDNP	3208

[illegible]


```

OY 2570 DKEEKAARRKREEVEQKRQNATKLSAL-----LFKHKEQDRAEITLKRALLDK 2622
Db 4124 -VEPFPAMSAOKKSESVYIRPTPTTGLAVISANTVGSILTEEMLIKISQPKODELIRQ 4182
OY 2623 DLQIEVQ-----EELKRDLKTKK-----EKDLMQLAQTAAACPPYVLPAPPPAP 2670
Db 4183 DSK-EVDSQDWSAKKEVNIQSVIKKLDTPLASKQAKRAVEQAL-----AP 4226
OY 2671 PSSPPPPGVQHTGLSTPTLPVASQKRKREEKSSSKSKKKM 2715
Db 4227 APIPNQPGNQ--SMAQETALPTTSSVNNSDHDEDETEFRQL 4269

RESULT 14
ID ABB66811 standard; Protein; 6815 AA.
XX ABB66811;
AC ABB66811;
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 27225.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX MPI: 2001-656860/75.
XX DR N-PSDB; ABLI0914.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure: SEQ ID NO 27225; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI57737-ABLI72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6815 AA;

Query Match 3.8%; Score 567; DB 22; Length 6815;
Best Local Similarity 17.1%; Pred. No. 2.6e-19;
Matches 609; Conservative 562; Mismatches 1238; Indels 1150; Gaps 145;

OY 2 VSEEEEDGGDAEEFQDSDDEDEMEER--DDDSDDYPE----- 39
Db 1210 VTELODEPQVEYETFRMS-DIEYVKLTLPSPMSDDKPTKRLIRHKGDEVOYLEIVESV 1268
OY 40 --MEDDDDAASYC-----ESSFRSHSYSSYTPGRKKRPVHHPRSPILFEKDIPLPE 89

```

Db	1269	VARVEDIKDMPADVLAEDEKIR-----	KKRVRAKDS-----	KOYEIIE	131	
Qy	90	FKSSEDLMVNREHIMNVAILEVLNF-----	GVULSLSPRREDFCAALY	136		
Db	1315	TEKPOD---IDPEARVITTT---EVSOGTIDGRABSTEARKKSVRRVKKREKKEIFVNIY		137		
Qy	137	SOEOCTMAEMHVULKAVLREEDPTSWTGTGPADLKDSVNSTLYFIDGTMWEVLVYCE		196		
Db	1371	BEARPLDHSVSELY-----	EDVLATTPRESSEKDI-----	PSFTTYVE	1408	
Qy	197	SUKEYHVLRYO-EAEDYRYGVENKIKV---IQFLVDQFLTNIAREELMSEGVIOYD		252		
Db	1409	-DEIVNPVLPREKICTVDVVRPKKKKKKIDONKKIKISEEPPTSEDSTIEEYTPKLE		1467		
Qy	253	HORVCHKIGDLLCSETGSAVYHLCVAPRYLEBYEDBMOCEYVA-----	H	298		
Db	1468	HD-----	EDLOTDESYDVANDSLPKSKKSTYKKOKE	1499		
Qy	299	KVVG-VTDGVAEIOKNRPYIRHEPIG---YDRSHRKWYFNRLILIEDTENENEKKIY		354		
Db	1500	SLEPGISLYTIRIETPT---EPTIKETIEDKEVYRVYINKRI-----	KKKA---	1544		
Qy	335	YSTKVOAELIDCDKDYWEALCKLLEKREELIHRMDITEDLTKANGSNKSFILAAN		414		
Db	1545	-GPKYEYILEVETIETEDNNPEADV-----	VIRIETPTSIDSKPOEDHKI-----	1587		
Qy	415	EBILESTAKKGDIDNV-----	KSPREYEROKN-----	ETENDSDAKKNE-----	456	
Db	1588	-QVQEKPKRESIDNTIOKLIDELIQVDHKEKAVLETIESPESKAKRIKHHKTE		1646		
Qy	457	-----EFEDQSLSEKSDDKTPD-----	DDPEQSGSEPTVGVGKGSVSAN	497		
Db	1647	VIDGPIVIEVTIOETETDEDKPKPBVYLKELIDHNEAEAKBEQVEITET-KAIDP		1708		
Qy	498	IGDNTT-----NATSEETSPSEGRPVGCISE--TPDSSNAEKVASEL-----	PQ	542		
Db	1706	LESEVVAITEDEQOEVLPAOEKKPIKOKKILKPEADVNYVVKVLEELTEPTOETIPE		1765		
Qy	543	DYPERNNKTCSSTMSATITSIQNLNENSSSELNSSOSESKADDPENERE---SH	59	599		
Db	1766	DADDPQVIEDISENNQVOI---IEEDGT---PKOYEIKKKVSKHBPQOVEI		1817		
Qy	600	TPVISOEIVDGFSEKSTGELSESPAGGASGSTRITRLNPNPSQLSKSOQVAA		659		
Db	1818	TETRSDEPLAEVITVELTEB---GLNMDI-----VIRQEKIYKKKKKIKLPPDISQY		1867		
Qy	660	AHEANKLEBKSG-----	EVLVNSQGEISIR--LSTKKEVYMK--GNIN	699		
Db	1868	VIRVLEENEPQMASTEPKPIIEDIASIEIIVPTEEDGTITKEVEVKKVSRKOGTQNO		1927		
Qy	700	YFKLGQBEKYAVYHNOYSTSPFLNKHQHEDDKRHLAHFCLLPRAEFKWNOSVGS		759		
Db	1928	VFEI-----	TEKTSDEPLAEVITIELSD-----	KS	1954	
Qy	760	KVLT-----	STRLTITOLENNI-----	PSFPHPNASHRAMIKAYOMC	801	
Db	1955	QEVITLPEKKPIKTIKLKPEDEVESVYVAVLEEFCEPOSFESPEPTBEAHETKT--KT		2012		
Qy	802	SKPREPALALILECAVPRVVMPLP---IWREFLGHTRLHRMTSIEREKKVYKKERKQ		857		
Db	2013	KKPK-----	KPIYKAPENVILIEEMAPETVIENTIENVEIGEEVQVQVTKTKLK	2059		
Qy	858	EFEETMOQATVKTTPVKKQVKKQKQGEERYVGYGSMWISKTHYRVRVPLBENTVYN		917		
Db	2060	KKEPKRELLIEIKETY-----		2075		
Qy	918	YKSLSEGKNNMDNMDESQKRCSCSPKKIKITEDPSEKDEVKGSADAKGADONEMDISK		977		
Db	2076	-----BENKREGD-----	IEITTELPEGSDASD--DOPIYIVQK	2110		
Qy	978	ITERK-----	DOQVCKLILDSDSK--PCKE-BEPEVDDOMKFKESHVNOCESSQVAVN--	1027		

Db 2111 IKKKPKKDDIKYIIOOLIEOITKTPLEEEPEEMDSKKPKKVKSHNKTIEVIDGL 2170
 QY 1028 -----VEGEHLRT-----SYKKTKSSKLDGLERRIKOF 1058
 Db 2171 PVTHFNEVDIYSEPEDEMERTLDEIKERIPOLPDDSSKYLVINISPEFEAKPIKOP 2230
 QY 1059 TLEKORLEK-----IKLEGGIKGIGKTSNKKLSSEPVITKAKECQSDSM 1107
 Db 2231 TODPIKKEPKKKKKKVEYVPSLEAFDHVAVVSEPTLEGTVEKVVKKRVSRRKGSK 2290
 QY 1108 RO--EOSPNNNDOPEDLIGCSDSSVLMSPSHTKNLYKRDVL--DDV----- 1157
 Db 2291 DHFIEITTSSEDPRAEVLVVEVLSDEPDSSEKPKHERKIVKPKQKDKDVEEYIIN 2350
 QY 1158 -----SIRSEPKCPKONSI-----END-----IEEKVSDLA 1184
 Db 2351 IIEFIOPIPVGLADEVKQKETEKKPKSPITTYIATBEDNDNNYDALVKEDLDOP 2410
 QY 1185 SRGOEPTKSKTGNDFTIDS-----KLASADDIGTLICKNKKPLIOESDTIV 1233
 Db 2411 ERALEKSSPLEYTVISEEESVGEOKOPKPKKISKPKSI-----KOPSVDKSPDYL 2463
 QY 1234 S--SSKALSSVPKSTNDRAT-----PLSRAMFE-----GKLGDSNSNTLESS 1280
 Db 2464 NVISEESIIDEPIEDYVTEAAEKPSEEPTEKVELETAVEREYTDGGETTKQSV 2523
 QY 1281 DTVSIO-----DSSEDMI-----VONSESISOEFTTR----- 1309
 Db 2524 TKRKIKKLVGKEEIIIEIVETKGTDPREYEVYVTEEOEKSEKAPKAKYKAKKI 2583
 QY 1310 -----EODVLEPLKCELVSGESTNCEDRLPVKGTENGKPKSO--OKKL 1354
 Db 2584 PKDDLQDYIOKLBIODIPKTELEKYEKIDLEBPVKMK--RKPICKVKVQSEOPKEETEPI 2642
 QY 1355 EEPVNCSDO-----IKLNTDJKNNENRESEK 1385
 Db 2643 EDKVEKEISESEVDSBPKITAVAKFEIPKPEKPEIYVLETVESKREPEEGKVR 2702
 QY 1386 GQRTSTFOINGKDKPKIYKGECLKEI--SESRYVSGNVEP-----KV 1427
 Db 2703 EKVYKTKKI--KONRGSEVVDIIVEIDDTNNSVITVTTVTEPPDQPSVKOKRT 2760
 QY 1428 NNKKIITPNDIKSLYKESAIRPFGND--VIMEDFENRNSSETK-----SHLSSD 1479
 Db 2761 KKKIKKDEVEFVKRV--IEEAPQEGSVDLVIEDFVKPSEKRRKKPIKDKH--TSVE 2817
 QY 1480 AEGVNRDLSLETPSTKE-----SDSTOTTPSPASCP--SNSVNOVEMETSEVKKYTS 1533
 Db 2818 EETPHEDEVLLESVPEDSPUSDLLITVDSVPJIEEPEKNVQIEDTKPEKRRKKRPS 2877
 QY 1534 SPITSEE-----ESNLNDFIDENGLPINKNE--NVNGESKRTKVIIT-- 1573
 Db 2878 AKLIEENVPEOTVEKPLEALHTSDLEKRPVQEFISISTKEEOKHHTPEKKSSKISSEQ 2937
 QY 1574 -----EYVMTSTVAESKTVIKVEGDKOVVSSTECAKSTVTTTTVTKLST 1624
 Db 2938 PKOPSTOYEIYSVEHDLKBEERKPTVOYVIOSETNVEEFD--DTGVKHKQVTTKRLMR 2996
 QY 1625 PSTGSGVDIISV-----KEOSKVVVTTTVVDSLTTT 1655
 Db 2997 PAGESGEIILEVNRDDQPEAEITIVVEPEVNODEKPREKPKTKRKVKDDIDHYIO-- 3054
 QY 1656 GGLVMTSMYKSEYSTDKYKLMFSRPPKTRSG--TALPSYRKPVTKSTKSIFV--LPND 1713
 Db 3055 --KLIELETKTELEKYEKIEFEPIYKDKPLDSPIDVLDESFKVOKDKKSRSTKVPNE 3112
 QY 1714 DL--KLARKGGR-----EVPYFNNAKPA--LDIMPY----- 1743
 Db 3113 ETPVOEYAKVNVVEEAPQEPRIPOILEYKPYEVAVKKEVITEDKRPVOEKTTRVILK 3172
 QY 1744 --PSRPTFGITMKYRQIYKSLAGVSLMLRLMLASLRMDMAKVPFGGSGSTTEISET 1801
 Db 3173 IGPEOETFKIT-----MIESEDNDSVTIV-----DEEPIASPOSIEEHPDOSKE 3219

QY 1802 EIT-----TEEIKRNDVGPGRIFREYCIKKI--CP-----IGVPEP----- 1838
 Db 3220 KLAEPKKTAVKVKKDDLSDI-----VKLIEEIPKVLDLEKYEKEMEKVKVLTVS 3272
 QY 1839 --KETPTORK--GLRSSALRPKRPETPKOTGPVILIEVVAEELWEIRAFERVE 1892
 Db 3273 DSIEEPPKPSKOPISVLPDTHKPKTKTPK-----TPKTE 3308
 QY 1893 KEKQAVBOAKKRLLEOK--TVIATISTSPSTSTISIAQVMMAPISGVTTGK 1951
 Db 3309 DTDOQVDEPETTVDTTIDELIPTQAPREDATQAITSQOEKSTQDDTQDTIOKT 3368
 QY 1952 V--LTKVSPATVTFQONKNEHOTFATVWOGOSNSGVQVOO--KVIGI--IPSSGTS 2006
 Db 3369 VKHKTKNDQTSVETSLSLPEVHKDYQISI--IHEELVEEOPKLEFRVIDEVAEV 3424
 QY 2007 QOTFTSFQPTATVYVIRPNISGSGTTSNSQVITGPQIRPMVY-----IRPPLQ 2056
 Db 3425 EES--QPIVEEV--EDEEPQATEEFTVEDVTKPKSKKKVKKTKDDELICKMLE 3477
 QY 2057 OSTIGKAIIRTPVWVQAPQOVMTQIIRGQVS-----TAVSAPNTVSST 2102
 Db 3478 QE--TEKTELEKYEKIEFDVPKLPKLEPALEPIKIERKQKPYVTIADATVPKTVLK 3536
 QY 2103 PGOKSLTSATS-----TSNIO----- 2118
 Db 3537 PSKKRKEAEELVQOLPKFRKAKRNVLEVYPPALIPKTTIGAIKONGELSRNIEEAE 3596
 QY 2119 -----SSASOP-----PRPQGO--VK 2133
 Db 3597 ILKFRPHKTKIKKIKIDLEKVELEKYEKYISSEEPPEKTPYKPKKAPKPEKQEDVVK 3656
 QY 2134 LTMQOLQVLQGHGNGGLTVVIOGOGOTGOLQIPGVV--LPGGQOOLMAAMPNG 2191
 Db 3657 LKL-----GKGGKKPEEAPRNVTLKINPOKPOEV----- 3687
 QY 2192 TVORFLPTLATATTAATTTTIVSTTAAGTGEOROSKLSF--OMQVODKTLPPAOSSSV 2250
 Db 3688 -----EEVELKOKKKEVEIYEEQKKKKDGEFV 3716
 QY 2251 GRAKAQPOFAPSAKPOPOFOPADPOVOTOPVOTOTTVSSHVSEAOPTHAQSKP 2310
 Db 3717 -----VEPPESEFDPREYVPLEDEQIIEHPEIPEKVKPKSKTYKPKDKSKSEPTIVS 3770
 QY 2311 QVAA-----OSOPSNVQSGSPVR--VQSPQTRIR--PSTQSLSPGQOSQOVQTTTS 2359
 Db 3771 EIVAGVPKEEBALPEODVKFKKPERDAPETDSEIKLRPVQASKDENPDQALVTPRAE 3830
 QY 2360 QPIPIO-----PHTSIQIPSOQOPQOPQOVOSTQTLSSGOTLNOVSSPSRP 2408
 Db 3831 EPIQOIELDKAIDEXKPKKK--PKVYQPKQGLAHEEPFVSVKEEALVDPK--- 3885
 QY 2409 QLOIQOPQOYIAVPQLOQOVLSQIOSOVVAOIQAOQSGVPQOIKLQPIQ--IQOS 2465
 Db 3886 --IEIKPKDVKKKPKKAPVSEV--VIEEPPREPEVPEIPEYKITTITVLEPE 3939
 QY 2466 SAVOTHOIQNVTVQASVQOQLOVRQOLRDQOQKQOQ--QIINNVTSPKLLIKYEII 2523
 Db 3940 DAPREHOVKVVIDFEROETTEEVIIEEKVVRKKKPKQOPEFEVTLKEPEEQIQDPV 3999
 QY 2524 OKOVMMKNNAVIEHLKOKSM-----TPAREBNQ-----RMIVCNOVMKYLIDKI 2569
 Db 4000 SAEISLP--IEEPQKPEYVELKTIQTTPEPNDVQJAVAEKKTGVKVKKEDKI 4055
 QY 2570 ---DKEEKQAARRKKEESVEOKRSROMATKLSALLFKHKEQILRAEILKRALLDKLO 2625
 Db 4056 VVVAEERKQVEBETIVE--VEKOEKKSSSEKPKSYEFKISETSQIE-----EKP 4103
 QY 2626 IEVQBELKROLKIKKEDQMLQATAVAAFCPPVTYVLPAPRAPPSPPPPPVQHTGL 2665
 Db 4104 IEVAEAPPEETPKVYKVAEKFDYSFTLKETDEEKVITVDDQPEE----- 4150


```

QY      2686 LSPPTLPLVASOKRRREENDSSKSKKKMISTSKET---KKDTKLTXCKRPVDESKF 2742
Db       4151 -EAPVAVVRKKRKRKEBAVEAFVMTEPKIVETSVELAKOKRTK*----KPKKDSEEA 4204
QY      2743 YTGDCRCOMWYHRCVGIILOSEAELIDEVY--CPGOSTEDATVTLPTEKDEGLKRY 2800
Db       4205 QL-----AIKVSEAPVAEEVFSEAPESKIVEEVLA---EKPKETTIRY 4248
QY      2801 LNSLOAHKNAWPELEVVDN-----DADPYGVIKE--PMDLATMEERY 2842
Db       4249 SESEP-----KEEPSVEQFTVKRKRPSTVFADEPATELVIKESKPFAVYTEDAH 4299
QY      2843 Q-RRYYEKLTFEYADMTKI 2860
Db       4300 KTKKKPKKYTLDVEAEELKI 4318

RESULT 15
ABB67502
ID      ABB67502 standard; Protein; 3257 AA.
XX
AC      ABB67502;
XX
DT      26-MAR-2002 (first entry)
XX
DE      Drosophila melanogaster polypeptide SEQ ID NO 29298.
XX
KW      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical.
XX      Drosophila melanogaster.
OS
PN      WO200171042-A2.
PD
XX      27-SEP-2001.
PF
PR      23-MAR-2001; 2001WO-US09231.
PR      23-MAR-2000; 2000US-191637P.
XX      11-JUL-2000; 2000US-0614150.
PA      (PEKE ) PE CORP NY.
PI
XX      Venter JC, Adams M, Li PWD, Myers EW;
DR      WPI: 2001-656860/75.
DR      N-PSDB; ABLL1605.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX
XX      Disclosure: SEQ ID NO 29298; 21pp + Sequence Listing; English.
XX
CC      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA
CC      sequences (ABLS7737-ABSL2072).
CC      (ABBS7737-ABBS2072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 3257 AA;

Query Match          3.8%; Score 562.5; DB 22; Length 3257;
Best Local Similarity 18.9%; Pred. No. 1.6e-19;
Matches 587; Conservative 445; Mismatches 1102; Indels 977; Gaps 135;

3 SEEDEEGDAETODSEDEDEDEEED-----DDSDYPE--E 39

```

[illegible]


```

Db 1081 A1ETD1E0YBEK--RSNRKRSRIRNEKFKETFDITLSDHLD----AKKAENASLEISM-- 1132
QY 980 EKKDQVKELLDSDSKP--CKEPEMDV----DMKTESHVNCQESSQOVVAVVSEGFHL 1034
Db 1133 -----RKCOTLETOQSDPYTAKNKRNSGRLSRKRS--VINAAS-- 1170
QY 1035 RSYKAKKTSSKLDGLERIKOFTLEKORLEK-----IKLEGGIKGIGKTSTNSSKN 1088
Db 1171 ----EKDKSPSALSOSTER--KOLLNENPSKKKKTQEOSGNKKEAVVGPJDKTETSSTN 1224
QY 1089 LSESPIYTAKEGCGSDSMRQEOSPMANNDQPEDLIQCGSDSVLNRMDPHTTNKLY 1148
Db 1225 IUDK-----KSNEFSDAM-----QPSDRL--NOKESATFKUSISSPKKIM 1265
QY 1149 PKDRLVDVS--IRSPETKCPKONSIEENDIEKESVDLASRGOEPTKSKTKGDFIDS 1205
Db 1266 DQDKDLALSKGDSNPTIRDTGEDSRQDTKKHQEND--TKHEEDSSKILKAN---IDET 1320
QY 1206 KLASADDIGTLCKNKKPLIOESDITVSSSKALSHSVKSTNDR-----DATPLS 1257
Db 1321 KSSSE-----KDAEPIISKDSOQ--DSAKPRL--SKRKSNNKRRKKNEKKNDSIAES 1367
QY 1258 RANDEFGKIGCOSE-----SNSTLENSDPTVS IQDSSEEDMIVONSIESISQFTRRQ 1311
Db 1368 ---DIEGGFOVNTETVQATCSTPESNKKDMVSDETNEE---PNLSETEIGRIRRGQ 1420
QY 1312 DVEVLEPLKCELVGSESTGNCEDRLPVKGTENGKKRPSQOKLEBRPVNKCSDQIKLNT 1371
Db 1421 AFHIEHPKDDHITTPQN-----ENOSIACVNFEPKOYPLPESVSDIRPKMIPKTYLMT 1475
QY 1372 TDKKNNENRESEKKGQRTSTFOINGKDNKPKYIKGECLEKIEISRVASVNEPKVNNIN 1431
Db 1476 KNKTS-----LLASSEDPIVLEPO--KLITTSK--GDSNDPIDNAN 1513
QY 1432 KIIPENDIKSLYKESAIRPFIINGDIVIMEDFNERNSEFMSHLLSSDAGNRDSEJTL 1491
Db 1514 NL-----ETSTQDPKEHEFS--DOTTDSNII 1540
QY 1492 PS-----TKESDSTQTT--TPSASCPE-----NSVNOVED--ME 1522
Db 1541 PSCTKKSQIVFTPTPKSSQOTKNSFITTPNRS--PKSKRNYSKAKRLDNEFEESQNASE 1599
QY 1523 IETSEYKATVSSPTITEESNLSNDPIDENGDLPIKNENY--NGESKRKVIT----- 1573
Db 1600 SSASKYQKELRPTASCRIKRLVLIKRTPTSSLPNSRKSIFFKTPAKSKRLTKILESMEK 1659
QY 1574 -----EYTMSTVATESKTVIVKKGDKQTVVSTENCAKST--VTTTTTVTK 1621
Db 1660 TFSREPSVSLGEVNPDSPPVAASVAVLHESDRDLESNEIPEEVEDETEEASAEJTDNK 1719
QY 1622 LSTPSTGGSVDLISVKEOSKTVVTTVTDSLTGTLVTSMTSKEYSTRQV--KLIM 1679
Db 1720 LKKEDDHELEVNDICAKKNPIT--DDSTKASSNKSITSDSVLOE--TKDLSNLSLIN 1774
QY 1680 FSR-----PKTNSGTLAPSTYKRFVTKSKSIFVLPNNDIKUL--ARKGIREVPYNNYA 1734
Db 1775 ATOGEDTPIKELTEEVVNNKKTVEDESKKOEI-----LKLEDPNAALEEDTASTAKA 1827
QY 1735 KPALDIWYPSRPRTGCIWRYRLQTVKSLAGSLMLRLMLASLRMDDMAAKP--FGGS 1793
Db 1828 AEBMDLY-----IKESNNKSVLAEPEDVTODEELAOQSPINSSSE 1868
QY 1794 TETETSEITITTELIRKRDVGPYIGREFYCIKRICPIGVPEPTEPPOKGLRSS 1853
Db 1869 TTSYVDDEPPESTSSVVK-----SLRK-----REABSSQD---EAA 1902
QY 1854 LRPKRPETPKO--TGPVITETWAELELMEIRAFAEVKE--KAQAVEQAKKRLQEQ 1910
Db 1903 KKKORQDYKSLTG-----KKEQVKPARRROLAEVEERPELKRKTESEAKSTVOGR 1954
QY 1911 KPTVLA-----TSTSPTSST--TSTISPAQKVAVAPISGVTGTGKMLVTK---VG 1958
Db 1955 YSIIQNETIMSTTAPIRETNREASTSPSARK-----SAVQBAKHVETTKHIIIG 2006

```

```

QY 1959 SPATVTFQONKNFHQFATWVWOGQNSGVVOYQOQVGLIIPSTSTGSOQFT----- 2011
Db 2007 PPG-----KLLH-----SDSPAEVKKPMVQVLLSTLSLQKPSFTLDDGSP 2049
QY 2012 -----SFOPTATVYTRNT-----SGSGGTSNSOYIITGPQIRPQMTVYRTLQOSTLGKAI 2064
Db 2050 KIRKSLKKSIADENIDGDOSIFSSSVLNKNTSVVAPBRKNIVSVLSQSKDQVETPAAS 2109
QY 2065 IRTPVWVQAPQOVMTQ-----IIRGQPVST-----AVSAPNTVSS 2101
Db 2110 SETPILTK--KEKKTQKSTKKEGKNKTESKKSLVQOPQMKTKQSEAVGPKILN- 2165
QY 2102 TPQOKSLTASTSNIQSSASQPPRPOGOVKLTMAQLTQLTGHHGNOGLTVVI--Q 2157
Db 2166 ----KYLKSETESSR--KTIVSTVGRKQIGL-----EVLKPPSRKSEELVEAISKKQ 2215
QY 2158 GQCGTQGLQLOLIFQGYVLDPG-----GQOLQOAMPNGVYQRFLETPLATATTASTTTT 2213
Db 2216 SQVQRLSKIDGRKSEGTSLPQDPVSKSETALKALPK-----ETEFVYODA 2261
QY 2214 TVSTTAAGTGEORQSKLSPQMOYHODKTLPPAOSSSGPAKAPQRTAPQAPOPOTO-- 2271
Db 2262 EIKMSKGRGHQN-----AVAKNTKEOP-----KSKPKTEVRSLOABATELM 2304
QY 2272 ----POSPAPEVQOPEVOTQTTVSSSHVPSAOPTHAOSKPOVAASOPOSNVQOS 2327
Db 2305 DMSDQSDVSDIATPPESGGIFNVPGHM-----TFAISNNSIAPTTPMDSOQ----- 2354
QY 2328 VVQSPSQTRIRPSTPSLSPGGOOSOVOTTTSDPIPIOPHTSLQIPSOQOPOSPOVOSS 2387
Db 2355 ----RNASKERFTTP-----VSDQKKPIRESQTLKRRARGRNQPLVSKR 2395
QY 2388 TQTLSSGQTLQNVQSVSSPSRPOLOIQOPPOVYLAVPQLQOQVYVLSQIQSQVAAQIQAOQ 2447
Db 2396 ----KAGEADGTAIVNPKPREDEEDHPQ-----QN 2424
QY 2448 SGVPOQIKLOLPIQIOSSAVQTHQIONVTVQAAVQEOLOQ-----RVQO-----LR 2495
Db 2425 DHVQESAFAFVPVKITPAASV--IPQVVRSGVNTVPQUMISPRKLCVKLNRRPYNKLWR 2480
QY 2496 DQOQKKQOQIEINVTPSKLL-----IKVEITQOVYMKHNAVIEHLKOKKS 2543
Db 2481 STQERNQEOGSRNV--TSLPLGETSETDGAASMSESILOSQV--OSEPAIOPL----- 2532
QY 2544 MTPAERENQRMVQVQVMYIILDKIDKEKQAAKKRRESEV-----EQKRSQNA 2595
Db 2533 --PASQDSCTL-----QASDLKIRESSQOLAPIAAYDSPAANDSS 2571
QY 2596 TKLSALIFKHKEQLRAEILKRALLDKLOIEVQOELKRLDKIKKEKDLQOLAATAVAA 2655
Db 2572 TSPALDIAPASQATAKATLMTALCPSTEKHLRDEPTLLESSKVAVEPOKQIQTQAKCLPV 2631
QY 2656 PCPPV-----TPVLPAP--PAPPPSPPPPEGVQHTG--LSTPLLPVVASOKRRR 2700
Db 2632 PIPEVASEPEDIMDEHSPNPMVAAAPATPQPHAITEDAGDGTQIOVNTLGVSTSRPL 2691
QY 2701 EEEK--DSSSKSKKKKMIISTSKETKKDKTLYCICITPYVESKFYIGCDRC 2749
Db 2692 ELHSIPASDPGPNPNAIGQTKMYSFLYPRY--KQSYDD---VGLDQC 2735

```

Search completed: November 20, 2002, 16:25:02
 Job time : 130.106 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 2002, 16:18:05 ; Search time 70.894 Seconds

(without alignments)
5227.100 Million cell updates/sec

Title: US-09-698-295-10

Perfect score: 14333

Sequence: 1 MVSEEEEDDGAETDSE.....KIKGFKASRSHNNKLTASTAS 2781

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues 908470

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14333	100.0	2781	21 AAY57453	Human transcriptio
2	14170	98.9	2907	21 AAY57452	Human transcriptio
3	2695.5	18.8	557	22 AAB94078	Human protein sequ
4	1537.5	10.9	976	22 ABB66735	Drosophila melanog
5	1474	10.3	328	21 AAB54364	Human pancreatic c
6	982.5	6.9	238	20 AAY04323	Human secreted pro
7	981	6.8	237	22 AAU19446	Human diagnostic a
8	789	5.5	149	23 AAU78845	SCAN/KRAB protein
9	781	5.4	246	22 AAU31944	Novel human secret
10	672.5	4.7	997	22 ABB62887	Drosophila melanog

11	572.5	4.0	2768	22 ABB68397	Drosophila melanog
12	560	3.9	5533	22 ABB65772	Drosophila melanog
13	560	3.9	5560	22 ABB71160	Drosophila melanog
14	549.5	3.8	6815	22 ABB66811	Drosophila melanog
15	534	3.7	3080	22 ABB64877	Drosophila melanog
16	529.5	3.7	3257	22 ABB67502	Drosophila melanog
17	508.5	3.5	2897	22 ABB58514	Drosophila melanog
18	441.5	3.1	4498	22 ABB58595	Drosophila melanog
19	434.5	3.0	2977	22 ABB69480	Drosophila melanog
20	432.5	3.0	1878	22 AAM40239	Human polypeptide
21	432.5	3.0	2137	23 ABB39618	Staphylococcus epi
22	432.5	3.0	3266	21 ABB42491	Human ORF225
23	430	3.0	1795	22 ABB6806	Drosophila melanog
24	430	3.0	2586	22 ABB68078	Drosophila melanog
25	429	3.0	3726	22 ABB63947	Human protein sequ
26	428	3.0	1727	22 ABB95554	Human transcriptio
27	423	3.0	2285	22 ABB63057	Drosophila melanog
28	418	2.9	4019	22 ABB13839	Human lung tumour
29	416.5	2.9	2665	22 ABB28314	Human peptide #95
30	416.5	2.9	2665	22 ABB33490	Peptide #96 encod
31	416.5	2.9	2665	22 ABB18950	Protein #99 encod
32	416.5	2.9	2665	22 AAM54270	Human brain expres
33	416.5	2.9	2665	22 AAM66665	Human bone marrow
34	416.5	2.9	2665	22 AAM14533	Peptide #97 encod
35	416.5	2.9	2665	22 AAM26950	Peptide #98 encod
36	416.5	2.9	2665	22 AAM02259	Peptide #941 encod
37	416.5	2.9	2665	23 ABB36319	Human peptide enco
38	410.5	2.9	1878	19 AAM81170	Human Bat2-alpha p
39	410	2.9	3596	23 ABB40235	Staphylococcus epi
40	409.5	2.9	2951	22 ABB60291	Drosophila melanog
41	404.5	2.8	3256	21 AAY50976	Human cell cycle p
42	403.5	2.8	1798	22 ABB71695	Drosophila melanog
43	402.5	2.8	1589	22 AAM42025	Human polypeptide
44	401	2.8	2344	22 AAU37120	Staphylococcus aur
45	400	2.8	2346	22 ABB63519	Drosophila melanog

ALIGNMENTS

RESULT 1	
AA157453	
ID	AA157453 standard; Protein; 2781 AA.
XX	
AC	AA157453;
XX	
DT	22-FEB-2000 (first entry)
XX	
DE	Human transcriptional regulatory factor SEQ ID NO:10.
XX	
KW	Human; transcriptional regulatory factor; TCOAL; BLAST detection;
KW	bromo-domain; cell proliferation; cancer.
XX	
OS	Homo sapiens.
XX	
PN	W09957143-A1.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99WO-TP02340.
XX	
PR	30-APR-1998; 98JP-0137631.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	
PI	Jones MH;
XX	
DR	WPI: 2000-052940/04.
DR	N-PSDB; AA39033.
XX	
PT	Transcriptional regulatory factor containing a bromo domain and gene
PT	TCOAL encoding it
XX	

PS Claim 1; Page 139-151; 154pp: Japanese.

XX The present sequence represents a human transcriptional regulatory factor containing a bromo domain. The factor interacts with proteins involved in the chromatin-mediated transcription regulatory mechanism. It binds to hSNF2L, hSNF2L and NCoA-62/Skip. It can be used for screening compounds binding to it and acting as agonists or antagonists, which are potentially useful for the treatment and prevention of cancer and other cell proliferation disorders.

XX Sequence 2781 AA:

Query Match 100.0%; Score 14333; DB 21; Length 2781;

Best Local Similarity 100.0%; Pred. No. 0; Matches 2781; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVSSEEEEDGDAEETODSDEDEDEEDDDSDYPEMEDDDDASTCTESSFSHST 60
DB 1 MVSEEEEDGDAEETODSDEDEDEEDDDSDYPEMEDDDDASTCTESSFSHST 60
QY 61 YSSTPGRRKRVHRRSPILIEKDIPLFPKSSSEDLAVNEHIMVIAIYVLRNGTV 120
DB 61 YSSTPGRRKRVHRRSPILIEKDIPLFPKSSSEDLAVNEHIMVIAIYVLRNGTV 120
QY 121 LRLSPFFEDFCALVSOEECTLMAEMHVLLKAVLREEDTSNTTEGPADLKDSVNSTLY 180
DB 121 LRLSPFFEDFCALVSOEECTLMAEMHVLLKAVLREEDTSNTTEGPADLKDSVNSTLY 180
QY 181 FIDGMTPYVLRYCESDKEYHNVLRYOEAEDVPYGVENKIKVLOPLVDFLTNTARE 240
DB 181 FIDGMTPYVLRYCESDKEYHNVLRYOEAEDVPYGVENKIKVLOPLVDFLTNTARE 240
QY 241 ELMSEGVIOYDDHGRVCHKIGDILCCETCSAVYHLECVKPLLEVPDEMOCVCVAHKY 300
DB 241 ELMSEGVIOYDDHGRVCHKIGDILCCETCSAVYHLECVKPLLEVPDEMOCVCVAHKY 300
QY 301 PGVTDVCAETIOKNNPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIMYYSTKVQ 360
DB 301 PGVTDVCAETIOKNNPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIMYYSTKVQ 360
QY 361 LAELIDLDLDYMAELCKLIEEMREIHRHMDITTEDLTJKARSNKSFLLAANEELLES 420
DB 361 LAELIDLDLDYMAELCKLIEEMREIHRHMDITTEDLTJKARSNKSFLLAANEELLES 420
QY 421 IRAKGGIDWVKSPEETEEKNETENDSKDAEKREFEFQOSLEKSDDKTPDDPPEQCK 480
DB 421 IRAKGGIDWVKSPEETEEKNETENDSKDAEKREFEFQOSLEKSDDKTPDDPPEQCK 480
QY 481 SEVGDFKSEKSNGLSESPGAGSGASTRIITRLRNPDSKLSQKSOVAAAAHEANKL 540
DB 481 SEVGDFKSEKSNGLSESPGAGSGASTRIITRLRNPDSKLSQKSOVAAAAHEANKL 540
QY 541 FKEKEVLVNVSGEISRLSTKKEVIMKGINNNFKLGQSGKRVYNNQYSTNFALNKH 600
DB 541 FKEKEVLVNVSGEISRLSTKKEVIMKGINNNFKLGQSGKRVYNNQYSTNFALNKH 600
QY 601 QHREDHDKRRHLAFCFLTAGEFKWNGSVGSKVLTISTRLITITOLENNIPSSFLHPN 660
DB 601 QHREDHDKRRHLAFCFLTAGEFKWNGSVGSKVLTISTRLITITOLENNIPSSFLHPN 660
QY 661 WASHRAWIKAVOMCSRPREFALALILECAVKKPVMLPIWREFLGHTRLRHMTSIEREE 720
DB 661 WASHRAWIKAVOMCSRPREFALALILECAVKKPVMLPIWREFLGHTRLRHMTSIEREE 720
QY 721 KEKVKKKKKOEBEETMOQATWVYTPPVKHQVWKQGEERYVVGSGWMSISTHYRRF 780
DB 721 KEKVKKKKKOEBEETMOQATWVYTPPVKHQVWKQGEERYVVGSGWMSISTHYRRF 780
QY 781 VPKLPJGNTNNYRKSLGTRKNNMDENDESDRKCSRPKKIKTEPSEKDEVGSDAAK 840
DB 781 VPKLPJGNTNNYRKSLGTRKNNMDENDESDRKCSRPKKIKTEPSEKDEVGSDAAK 840
QY 841 GADONEMDISKITEKKDQVKELLSDSDKPCKEEPMEDDDMKTESHVNCQESSQVYV 900

DB 841 GADONEMDISKITEKKDQVKELLSDSDKPCKEEPMEDDDMKTESHVNCQESSQVYV 900
QY 901 NVSBGFRLRTSYKKTKTKSSKLDGLERRIKOPLLEKQRLKIKLGGCIKIGIKTSTNSS 960
DB 901 NVSBGFRLRTSYKKTKTKSSKLDGLERRIKOPLLEKQRLKIKLGGCIKIGIKTSTNSS 960
QY 961 KNISESPVITRAKEGCOSDMROEOSPANNNDPREDLIQCSODSSVLMSDPSSHNTNK 1020
DB 961 KNISESPVITRAKEGCOSDMROEOSPANNNDPREDLIQCSODSSVLMSDPSSHNTNK 1020
QY 1021 LYPKRDVLDVSIKSPETKCPKONSIENTIEEKVSDLASGOEPYKSKTGNDFTIDSK 1080
DB 1021 LYPKRDVLDVSIKSPETKCPKONSIENTIEEKVSDLASGOEPYKSKTGNDFTIDSK 1080
QY 1081 LASADDTGLIKNNKKRLIOEESDTIVSSSKSAHSSVPRKSTNDROATPLSRAMDEGKL 1140
DB 1081 LASADDTGLIKNNKKRLIOEESDTIVSSSKSAHSSVPRKSTNDROATPLSRAMDEGKL 1140
QY 1141 GCDSESNSTLENSDVTYSIODSSPEDMIYONSNEISIEQFRTREODEVLEPLKCELVSG 1200
DB 1141 GCDSESNSTLENSDVTYSIODSSPEDMIYONSNEISIEQFRTREODEVLEPLKCELVSG 1200
QY 1201 ESTGNCEDRLPVKGTENGKKPSQOKKLEBRPVNKCSDQIKLNTDKNENRESEKKG 1260
DB 1201 ESTGNCEDRLPVKGTENGKKPSQOKKLEBRPVNKCSDQIKLNTDKNENRESEKKG 1260
QY 1261 QRTFOINKDKPKTYLGECLKETSSESRVYSGNNEPVNNINKIIPNDIKSLTVKE 1320
DB 1261 QRTFOINKDKPKTYLGECLKETSSESRVYSGNNEPVNNINKIIPNDIKSLTVKE 1320
QY 1321 SAIRPFINGDVIMEDFENRNSSETKSHLSSPSAEGVNRBSLETLPSKTESDSTQTTTPS 1380
DB 1321 SAIRPFINGDVIMEDFENRNSSETKSHLSSPSAEGVNRBSLETLPSKTESDSTQTTTPS 1380
QY 1381 ASCPESNVNOVEDMELETSEVKKVTSSPTTSEBSNLSDFIDENCLPINKENNNGES 1440
DB 1381 ASCPESNVNOVEDMELETSEVKKVTSSPTTSEBSNLSDFIDENCLPINKENNNGES 1440
QY 1441 KRKTIVTEVTMTSTVATESKTYAKVKGDKQFVASTENCASKSTVTTTTYTKLSTPS 1500
DB 1441 KRKTIVTEVTMTSTVATESKTYAKVKGDKQFVASTENCASKSTVTTTTYTKLSTPS 1500
QY 1501 TGGSVDLISVKEOSKTVVTTVTDSLTGTTGTLVTSMTVSKYSTRDYKLMKFSRPKPT 1560
DB 1501 TGGSVDLISVKEOSKTVVTTVTDSLTGTTGTLVTSMTVSKYSTRDYKLMKFSRPKPT 1560
QY 1561 RSGTALPSTYKRFYTKSTKKSIFVLPNDLKLARKGIREVPFNNYNAKALDIWPPSP 1620
DB 1561 RSGTALPSTYKRFYTKSTKKSIFVLPNDLKLARKGIREVPFNNYNAKALDIWPPSP 1620
QY 1621 RPFPGITWRYRLQTVKSLACVSLMLRLMASLAMDMAAVPPGGGSTRTSETETITTT 1680
DB 1621 RPFPGITWRYRLQTVKSLACVSLMLRLMASLAMDMAAVPPGGGSTRTSETETITTT 1680
QY 1681 EIKRRDVGYGIRFEYCIRKIIICPIGPVPTPKETPTPQKGLRSSALRPKRPETPKQNG 1740
DB 1681 EIKRRDVGYGIRFEYCIRKIIICPIGPVPTPKETPTPQKGLRSSALRPKRPETPKQNG 1740
QY 1741 PVIIETVVAEELELMIRAFARVEKEKAQAEQAKKLEQOKPVIVATSTTSPSSST 1800
DB 1741 PVIIETVVAEELELMIRAFARVEKEKAQAEQAKKLEQOKPVIVATSTTSPSSST 1800
QY 1801 TSTISPAQKVMVAPISGSVTTGTRMVLTTVGSFATVFOONKNFOTFTATWKQGSNS 1860
DB 1801 TSTISPAQKVMVAPISGSVTTGTRMVLTTVGSFATVFOONKNFOTFTATWKQGSNS 1860
QY 1861 GVVQVQKVLGIIIPSSGTSGOFTSFQPRATVYIRPNTSGSGGTTSSNSQVITGPQIRP 1920
DB 1861 GVVQVQKVLGIIIPSSGTSGOFTSFQPRATVYIRPNTSGSGGTTSSNSQVITGPQIRP 1920
QY 1921 GMTVIRTPLOOSTLQKAIIRTPVWVOPGAQOQVWTOIIRQOPVSTAVSAPVYVSTIPGOK 1980
DB 1921 GMTVIRTPLOOSTLQKAIIRTPVWVOPGAQOQVWTOIIRQOPVSTAVSAPVYVSTIPGOK 1980


```
Db 1921 GMTVITPILQOSTLGKAILRTPVMPGAPQOVMTOIIRGOVSTAVSAPNTVSTPBGK 1980
Qy 1981 SLTSATSTSNIOSSASQPRPQOGVYKLTMAOLTQLTQGHGNOGLTVVIOGOGTTGOL 2040
Db 1982 SLTSATSTSNIOSSASQPRPQOGVYKLTMAOLTQLTQGHGNOGLTVVIOGOGTTGOL 2040
Qy 2041 QILIPGCVTVLPFGGQOLMOQAMPNGTVQRFLETPLATTTATTTTSTTTTSTTAAGTGEOR 2100
Db 2041 QILIPGCVTVLPFGGQOLMOQAMPNGTVQRFLETPLATTTATTTTSTTTTSTTAAGTGEOR 2100
Qy 2101 QSLKSPOMOVHODKTLPRPOSSSVGPAKAPOTAPPSARPOTOPOSPAPPEVOTQPEV 2160
Db 2101 QSLKSPOMOVHODKTLPRPOSSSVGPAKAPOTAPPSARPOTOPOSPAPPEVOTQPEV 2160
Qy 2161 QTOQTVSSHVPESEAPQTHAQSCKPQVAAQSOPQSNVQGSFPVAVQSPQTRIRPSTPSOL 2220
Db 2161 QTOQTVSSHVPESEAPQTHAQSCKPQVAAQSOPQSNVQGSFPVAVQSPQTRIRPSTPSOL 2220
Qy 2221 SPQOOSQOVOTTSQPIPIQPHSTLQIPSGQPOSPQOVOSSTQTLSSGQTLNVOVSPPS 2280
Db 2221 SPQOOSQOVOTTSQPIPIQPHSTLQIPSGQPOSPQOVOSSTQTLSSGQTLNVOVSPPS 2280
Qy 2281 RPOLQIQOPQOPIYAVPOLQOQOVVLQIOQOVVAVQIOAQOQSVPOQIKLQIPQIQOQS 2340
Db 2281 RPOLQIQOPQOPIYAVPOLQOQOVVLQIOQOVVAVQIOAQOQSVPOQIKLQIPQIQOQS 2340
Qy 2341 AVQTHQIQOVVTVQAASVQEQLOQVQOLRPDQOQKKQOQOIEIKREHTLQASNOSEIIOQK 2400
Db 2341 AVQTHQIQOVVTVQAASVQEQLOQVQOLRPDQOQKKQOQOIEIKREHTLQASNOSEIIOQK 2400
Qy 2401 VVKKHNAVIEHLKQKSKMTPAEREENORMIVCNOVMKYIILDKIDKEKQAKKRRKEESV 2460
Db 2401 VVKKHNAVIEHLKQKSKMTPAEREENORMIVCNOVMKYIILDKIDKEKQAKKRRKEESV 2460
Qy 2461 EOKRSQONTKLSALFLFKHKEOLRAELKKRALDLDDLOIEVOEELKRDLIKKEKDLQ 2520
Db 2461 EOKRSQONTKLSALFLFKHKEOLRAELKKRALDLDDLOIEVOEELKRDLIKKEKDLQ 2520
Qy 2521 LAQATVAAPPCRPVTPVLPAAPAPSPPPPGVQHTGLSTPTLPAVSQKRRKEEKDS 2580
Db 2521 LAQATVAAPPCRPVTPVLPAAPAPSPPPPGVQHTGLSTPTLPAVSQKRRKEEKDS 2580
Qy 2581 SSKSKKKKMIISTTSKTKKDTKLYCICKTPYDESKFYIGCDRCOMVHGRVGILOSEAE 2640
Db 2581 SSKSKKKKMIISTTSKTKKDTKLYCICKTPYDESKFYIGCDRCOMVHGRVGILOSEAE 2640
Qy 2641 LIDEVYCPOCQOSTEDAMTVLTLTEKDYEGIKRVLSIQAHKMAKMPLEFVNDNDADYY 2700
Db 2641 LIDEVYCPOCQOSTEDAMTVLTLTEKDYEGIKRVLSIQAHKMAKMPLEFVNDNDADYY 2700
Qy 2701 GVIKPEMDLATMEERVQQRKYETKLEFVADMTKIFDNCRYYNPSDSPFYQCAVLESFEV 2760
Db 2701 GVIKPEMDLATMEERVQQRKYETKLEFVADMTKIFDNCRYYNPSDSPFYQCAVLESFEV 2760
Qy 2761 QKLGFKASRSHNNKLOSTAS 2781
Db 2761 QKLGFKASRSHNNKLOSTAS 2781

RESULT 2
AA57452
ID AA57452 standard; Protein: 2907 AA.
XX
XX AA57452;
XX
XX 22-FEB-2000 (first entry)
XX
XX Human transcriptional regulatory factor SEQ ID NO:1.
XX
XX Human transcriptional regulatory factor; TCOAL; BLAST detection;
XX bromo-domain; cell proliferation; cancer.
OS Homo sapiens.
```

```
XX
PN WO957143-A1.
XX
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999. 99MO-JP02340.
PF
XX
XX 30-APR-1998. 98JP-0137631.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI
PI Jones MH;
XX
XX MPI: 2000-052940/04.
DR N-PSDB: AA239032.
XX
PT Transcriptional regulatory factor containing a bromo domain and gene
PT TCOAL encoding it -
XX
XX Claim 1; Page 48-68; 154pp; Japanese.
XX
CC The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulation mechanism. It binds
CC to hSNF2H, hSNF2L and NCoA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
XX
SQ Sequence 2907 AA;
XX
Query Match 98.9%; Score 14170; DB 21; Length 2907;
Best local Similarity 95.1%; Pred. No. 0;
Matches 2762; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
Qy 1 MVSSEEEEDGDAEFTQDSDEDEDEDEDDSDYPEEMEDDDDASTESSFNSHST 60
Db 1 MVSSEEEEDGDAEFTQDSDEDEDEDEDDSDYPEEMEDDDDASTESSFNSHST 60
Qy 61 YSTPGRRKRVHRPSPILTEEKDIPLEPRKSSDLMVNEHIMVATIEVLRNPGTV 120
Db 61 YSTPGRRKRVHRPSPILTEEKDIPLEPRKSSDLMVNEHIMVATIEVLRNPGTV 120
Qy 121 LRSPREFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSMTTGPADLKDSVNSTLY 180
Db 121 LRSPREFEDFCAALVSOEQCTLMAEMHVLLKAVLREEDTSMTTGPADLKDSVNSTLY 180
Qy 181 FIDGMTPEVLRVYCESDKETVHNVLPYQEAEDYRYGVEVKNKIKVLOFLVDQFLTTNARE 240
Db 181 FIDGMTPEVLRVYCESDKETVHNVLPYQEAEDYRYGVEVKNKIKVLOFLVDQFLTTNARE 240
Qy 241 ELMSEGVIOYDDHCRVCHKIGDLLCETCSAVVYHLCVYKRLFEVDEDEQCEVCAHKY 300
Db 241 ELMSEGVIOYDDHCRVCHKIGDLLCETCSAVVYHLCVYKRLFEVDEDEQCEVCAHKY 300
Qy 301 PGVTDCAVETQKKNPYIRHEPIGYDSRSRRKYWFNLNRLITEEDTENENKKIYYSTKYQ 360
Db 301 PGVTDCAVETQKKNPYIRHEPIGYDSRSRRKYWFNLNRLITEEDTENENKKIYYSTKYQ 360
Qy 361 LAELIDCLDKDYWEALCKLLEEMREIHRHMDITEDLTNKAQSNKSFLLAANEELLES 420
Db 361 LAELIDCLDKDYWEALCKLLEEMREIHRHMDITEDLTNKAQSNKSFLLAANEELLES 420
Qy 421 IRAKKGDIDNVKSPETETKDKNETENDSKAEKNREFEQSLSEKSDKTPDDPDQOK 480
Db 421 IRAKKGDIDNVKSPETETKDKNETENDSKAEKNREFEQSLSEKSDKTPDDPDQOK 480
Qy 481 SE----- 482
Db 481 SEPTVEGDKGNSVANLGDNTNATSEETSPSEGSPPVOCISLSETPDSSNMAKKVASEL 540
OS 483 ----- 482
```


Db 541 P0VPEEPNKTCESSENTSATTSTSI0PULENSNSSELNSQS0SEKAAADDPENGERSHT 600
QY 483 -----VDFKSEKNGELSESPGAGKASGSTRITTLRLRPNDKLSOLKSOQVAAA 534
Db 601 PVSIOEBIVDFSEKSTGSELSESPGAGKASGSTRITTLRLRNDKLSOLKSOQVAAA 660
QY 535 HEANKLEKEKEEVLVANS0GEISRLSTKKEVIMKGINNNFKLGOEKEVYVHNQYSTNS 594
Db 661 HEANKLEKEKEEVLVANS0GEISRLSTKKEVIMKGINNNFKLGOEKEVYVHNQYSTNS 720
QY 595 PALNKH0RHDHDKRRLAHKFCITPAGERKMGSVHGSVLTITSTLRTITTOLENNIPS 654
Db 721 PALNKH0RHDHDKRRLAHKFCITPAGERKMGSVHGSVLTITSTLRTITTOLENNIPS 780
QY 655 SFLHPNASHRAMNIRKAV0CSPREFALALITLCEAVKRVVMLPTIRELGHTRHMR 714
Db 781 SFLHPNASHRAMNIRKAV0CSPREFALALITLCEAVKRVVMLPTIRELGHTRHMR 840
QY 715 SIEREEKYKKEKKE0EETMOQATWVYTPFVKHQWK0KGEERYVNGYGMWSIK 774
Db 841 SIEREEKYKKEKKE0EETMOQATWVYTPFVKHQWK0KGEERYVNGYGMWSIK 900
QY 775 THYRFPKPLPGNTNVYKRSLEGTKNNDNDESKRCSNPKIKITIEPDEKEVK 834
Db 901 THYRFPKPLPGNTNVYKRSLEGTKNNDNDESKRCSNPKIKITIEPDEKEVK 960
QY 835 GSDAKADONEMDISKITEKKDODVKELDSDKPCKEEPMVDDDMKTESHVNC0ES 894
Db 961 GSDAKADONEMDISKITEKKDODVKELDSDKPCKEEPMVDDDMKTESHVNC0ES 1020
QY 895 S0VDVNVSEGFHLRTSYKKTSSKLDGLLERRIKOFTLEEKORLEKIKLEGIGIKIG 954
Db 1021 S0VDVNVSEGFHLRTSYKKTSSKLDGLLERRIKOFTLEEKORLEKIKLEGIGIKIG 1080
QY 955 TSTSNKLSLSPVITAKEGCOSMSKROSPYANNDOPEDLIOGOS0SSVLRMSDP 1014
Db 1081 TSTSNKLSLSPVITAKEGCOSMSKROSPYANNDOPEDLIOGOS0SSVLRMSDP 1140
QY 1015 SHTNKNLYPKDRVLDDVIRSPEKCPKONSIENTDIEEKYSDLASR0EPTKSTKNDP 1074
Db 1141 SHTNKNLYPKDRVLDDVIRSPEKCPKONSIENTDIEEKYSDLASR0EPTKSTKNDP 1200
QY 1075 FIDDSKIASADDTGTLCKNKKPLIOEESDTIYSSSKSALHSSVPKSTNDRDATPLSRAM 1134
Db 1201 FIDDSKIASADDTGTLCKNKKPLIOEESDTIYSSSKSALHSSVPKSTNDRDATPLSRAM 1260
QY 1135 DFEKLCGDSNSNTLENSSDTVSIO0SSEDMIVONSNEISIQ0FTR0ED0VVLPLK 1194
Db 1261 DFEKLCGDSNSNTLENSSDTVSIO0SSEDMIVONSNEISIQ0FTR0ED0VVLPLK 1320
QY 1195 CEIVSGESTGNCEDRLPVKGT0EANGKRP00KLEERPVKCSDQIKLKTNTD0KNNENR 1254
Db 1321 CEIVSGESTGNCEDRLPVKGT0EANGKRP00KLEERPVKCSDQIKLKTNTD0KNNENR 1380
QY 1255 ESEKKG0RTSTFOINGKDNKPKIYLKGECLKEISESRVUSGVNPKYNNINIKITPENDIK 1314
Db 1381 ESEKKG0RTSTFOINGKDNKPKIYLKGECLKEISESRVUSGVNPKYNNINIKITPENDIK 1440
QY 1315 SLVTKESAIRPFIINGDIYMEDEFERNSESETKSHLLSSDAAGNRDSELTLPSTKESDST 1374
Db 1441 SLVTKESAIRPFIINGDIYMEDEFERNSESETKSHLLSSDAAGNRDSELTLPSTKESDST 1500
QY 1375 0TTTPSASCPEBSNVNOVEDMEIETSEVKVYTSPIITSEBSNLSNDFIDENGLPIKNE 1434
Db 1501 0TTTPSASCPEBSNVNOVEDMEIETSEVKVYTSPIITSEBSNLSNDFIDENGLPIKNE 1560
QY 1435 NVN0ESKRKYITTEVTMTSTVATESTKVIYKEKGDKQTVVSSITENAKSVTTTTTIVT 1494
Db 1561 NVN0ESKRKYITTEVTMTSTVATESTKVIYKEKGDKQTVVSSITENAKSVTTTTTIVT 1620
QY 1495 KLSTPSTGGSVDLISVKEOSKTVVTTVTBDSLTGGLVTSMMVSKESYSTRDVKLAKF 1554
Db 1621 KLSTPSTGGSVDLISVKEOSKTVVTTVTBDSLTGGLVTSMMVSKESYSTRDVKLAKF 1680

QY 1555 SRPKTRSGTALPSYKREFTVKSTKSTIFVLPNDLKKLAKKGIR0EVPYENYNAKPALDI 1614
Db 1681 SRPKTRSGTALPSYKREFTVKSTKSTIFVLPNDLKKLAKKGIR0EVPYENYNAKPALDI 1740
QY 1615 WPYSPRPFPFGITWRYRRL0TVKSLAGVSLMLRLMLASLRMDMAKAVPPGSGSTR0ETSE 1674
Db 1741 WPYSPRPFPFGITWRYRRL0TVKSLAGVSLMLRLMLASLRMDMAKAVPPGSGSTR0ETSE 1800
QY 1675 TEITTTTILIKR0VGPFGIFEFYCIKIKIICPIGVPEPKTPTP0KGLRSSALRKRP 1734
Db 1801 TEITTTTILIKR0VGPFGIFEFYCIKIKIICPIGVPEPKTPTP0KGLRSSALRKRP 1860
QY 1735 TP0T0GVIIEETWVAE0ELMEIRAFABEVEKEKA0A0EQAKKRL00KRPVIATSTT 1794
Db 1861 TP0T0GVIIEETWVAE0ELMEIRAFABEVEKEKA0A0EQAKKRL00KRPVIATSTT 1920
QY 1795 SPTSSSTTSTISPA0KVVADISSVTTGTIMVLTITKGSFAVTF00KNF0T0FATWYK 1854
Db 1921 SPTSSSTTSTISPA0KVVADISSVTTGTIMVLTITKGSFAVTF00KNF0T0FATWYK 1980
QY 1855 0G0SNSGV0V00KVLGIIIPSTGTS0QFTSP0PRTATVTIRPNTSGSGGTTSN0VIT 1914
Db 1981 0G0SNSGV0V00KVLGIIIPSTGTS0QFTSP0PRTATVTIRPNTSGSGGTTSN0VIT 2040
QY 1915 GPOIRP0MTVIRTPLO0STLGKAIIRTPVWY0GAP0QVFT0IIR0QPVSTAVSAPNTVS 1974
Db 2041 GPOIRP0MTVIRTPLO0STLGKAIIRTPVWY0GAP0QVFT0IIR0QPVSTAVSAPNTVS 2100
QY 1975 STP0KSLTATSTSNIS0SASQPPRP0G0GVKLTMAQLT0L0V0HGNGGLTVVY0G0G 2034
Db 2101 STP0KSLTATSTSNIS0SASQPPRP0G0GVKLTMAQLT0L0V0HGNGGLTVVY0G0G 2160
QY 2035 0TTG0L0LIP0GVTLP0P0Q0LMA0AMPNGT0V0R0FLTPLATTA0STTTTIVSTTA 2094
Db 2161 0TTG0L0LIP0GVTLP0P0Q0LMA0AMPNGT0V0R0FLTPLATTA0STTTTIVSTTA 2220
QY 2095 GTG0R0RSKLSLPM0V0H0DKTLPPA0SSSVGPAKAPQTA0PSARPQTA0P0SPAP0EV 2154
Db 2221 GTG0R0RSKLSLPM0V0H0DKTLPPA0SSSVGPAKAPQTA0PSARPQTA0P0SPAP0EV 2280
QY 2155 0T0P0EV0T0TTVSSHVSEK0PHTHA0SSKPOVAA0S0P0SNV0G0S0VVR0QSP0RIRP 2214
Db 2281 0T0P0EV0T0TTVSSHVSEK0PHTHA0SSKPOVAA0S0P0SNV0G0S0VVR0QSP0RIRP 2340
QY 2215 STP0SLSPG00S0V0TTS0PIPI0PHTSL0IS0G0P0S0P0V0S0T0TLSSG0TLNV 2274
Db 2341 STP0SLSPG00S0V0TTS0PIPI0PHTSL0IS0G0P0S0P0V0S0T0TLSSG0TLNV 2400
QY 2275 SVS0SPSRP0L0I0Q0P0VYTAVP0L00Q0V0VLS0IG0VYVA0I0A0Q0SGVPO0IK0LPI 2334
Db 2401 SVS0SPSRP0L0I0Q0P0VYTAVP0L00Q0V0VLS0IG0VYVA0I0A0Q0SGVPO0IK0LPI 2460
QY 2335 0I00SSAV0TH0I0NVVYV0AASV0E0L0V00L0R00Q0KK0Q0I0INNTWSKILLIKV 2394
Db 2461 0I00SSAV0TH0I0NVVYV0AASV0E0L0V00L0R00Q0KK0Q0I0INNTWSKILLIKV 2520
QY 2395 EI10K0VVMKNNAVIEHLK0KKSMTPA0REB0N0MIVC0NMVYILDKIDKEE0QA0KRR 2454
Db 2521 EI10K0VVMKNNAVIEHLK0KKSMTPA0REB0N0MIVC0NMVYILDKIDKEE0QA0KRR 2580
QY 2455 KREESV0KRSK0NATRLSALLKFKH0LRAEILK0KALLDK0L0F0V0E0ELK0RDLKIK 2514
Db 2581 KREESV0KRSK0NATRLSALLKFKH0LRAEILK0KALLDK0L0F0V0E0ELK0RDLKIK 2640
QY 2515 EK0LM0LA0ATVAVAP0CPVTPVLPAPAP0SP0PP0G0H0GLSTPLLPVAS0KRR 2574
Db 2641 EK0LM0LA0ATVAVAP0CPVTPVLPAPAP0SP0PP0G0H0GLSTPLLPVAS0KRR 2700
QY 2575 EEEK0SSSKK0KMMISTSKETKKDTKLYCICKTPYDESKFYIG0DR0C0NMWYHGR0VGI 2634
Db 2701 EEEK0SSSKK0KMMISTSKETKKDTKLYCICKTPYDESKFYIG0DR0C0NMWYHGR0VGI 2760

0y	2635	LSAEALIDDEVVCCOOSTEDAMVVLPLETEKQEGALKRVLRLSLQAKKAMPLEPEVDN	2694
0y	2635	LSAEALIDDEVVCCOOSTEDAMVVLPLETEKQEGALKRVLRLSLQAKKAMPLEPEVDN	2694
Db	2761	LSSEALIDDEVVCCOOSTEDAMVVLPLETEKDEGLKRLRLSLQAKKAMPLEPEVDN	2820
0y	2695	DAPDYGVYIKPEPMLATMEERVQRYYEKLTPEFVADMTKLFQDNCRRYNPSDSPEYOCAEV	2754
Db	2821	DAPDYGVYIKPEPMLATMEERVQRYYEKLTPEFVADMTKLFQDNCRRYNPSDSPEYOCAEV	2880
0y	2755	LESFVQKLKGFKASRSHNNKLSQSTAS	2781
Db	2881	LESFVQKLKGFKASRSHNNKLSQSTAS	2907
RESULT 3			
ID	AAB94078	AAB94078 standard; Protein: 557 AA.	
AC	AAB94078;		
DT	26-JUN-2001	(first entry)	
XX	Human	Human protein sequence SEQ ID NO:14273.	
DE	Human	Human protein sequence SEQ ID NO:14273.	
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy.		
XX	Homo sapiens.		
OS	Homo sapiens.		
PN	EP1074617-A2.		
XX	07-FEB-2001.		
PD	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
PR	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
PI	WPI: 2001-318749/34.		
DR	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length cDNAs defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length cDNAs -		
XX	Claim 8; SEQ ID 14273; 2537pp + CD ROM; English.		
PS	The present invention describes primer sets for synthesizing 5602		
XX	full-length cDNAs defined in the specification. Where a primer set		
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		
CC	in gene therapy. The primers are useful for synthesizing polynucleotides		
CC	particularly full-length cDNAs. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length cDNAs. The primers allow obtaining of the full-length		
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to		
CC	AAH95983 represent human amino acid sequences; and AAH13629 to AAH13632		

CC		represent oligonucleotides, all of which are used in the exemplification
CC		of the present invention.
XX		
SQ	Sequence	557 AA;
	Query Match	18.8%; Score 2695.5; DB 22; Length 557;
	Best Local Similarity	79.0%; Pred. No. 1,6e-130;
	Matches	553; Conservative 3; Mismatches 1; Indels 143; Gaps
OY	1811	MVAPISGSVTGTMVLTTKVGSPATVTFPOONKFEHDTFATWWKOGSNGGVVOVKYL 1870
DB	1	MVAPISGSVTGTGMVLTTKVGSPATVTFPOONKSFHOTFTFWKOGSNGGVVOVKYL 60
OY	1871	GILPSSTGSTSOOTFTSFOPRTATVTIIPNTSGSGCTTNSNSOVIITGPDIRPGMTVIRTPDQ 1930
DB	61	GILPSTGTSGQTFTSRQPRTAVTITRPNTSGSGCTTNSNSOVIITGPDIRPGMTVIRTPDQ 120
OY	1931	OSTLGAKAIIPTPVWVQCAPQOVNTOIIRKCPYSTAVSANPTVSSITPGOKSLTSATSTSN 1990
DB	121	OSTLGAKAIITPPVMVQCAPQOVNTOIIRKCPYSTAVSANPTVSSITPGOKSLTSATSASN 180
OY	1991	IOSASASOPRRQOCQVLTMAQLTOLNOGHGNGGLNVIVIOGGQITTGQDLIPCGVTYL 2050
DB	181	IOSASASOPRRQOCQVLTMAQLTOLNOGHGNGGLNVIVIOGGQITTGQDLIPCGVTYL 240
OY	2051	PGPGQQLLQAAMPNGTVORLEFLTPLATTAATASTTTTTVSTAAGTGEORQSRLSPOMCV 2110
DB	241	PGPGQQLLQAAMPNGTVORLEFLTPLATTAATASTTTTTVSTAAGTGEORQSRLSPOMCV 300
OY	2111	HODKTLPRAQSSSVGPAKAOPOTAOPSARPOPOTOPOSPAPAEVOTQPEVOTQTVSSHV 2170
DB	301	HODKTLPRAQSSSVGPAEAOPOTAPSARPOPOTOPOSPAPAEVOTQPEVOTQTVSSHV 360
OY	2171	PSEKAPTTHAQSSKKQVAQAQSQPOSNNQGSGSVRKVSQRIRPSTRPSQLSPGOQSOYQT 2230
DB	361	PSEKAPTTHAQSSKKQVAQAQSQPOSNNQGSGSVRVQSPSORIRPSTRPSQLSPGOQSOYQT 420
OY	2231	TTSOPRIPTHTSIQIIPSQGPQSPQPOVSQSTQLSSGTLNQVSVSSPRLDIQQDPQ 2290
DB	421	TTSOPRIPTHTSIQIIPSQGPQSPQ----- 446
OY	2291	POVIAVPQLDQOVULSQIISOVVAAIQAAQSGVPQIKLRLPIQIOSSAVQTHQIDNV 2350
DB	447	----- 446
OY	2351	VTVDAASVQEDELQRYQQLRDQOKKKQOOIEIKREHTLAQSNOSSEIIQKVYCKHNNAVTE 2410
DB	447	-----QVYCKHNNAVTE 457
OY	2411	HLKOKKSMTPARERENORMIVCNOMKYLLDKIDKEKOAKRKRRESVEOKRSRONMT 2470
DB	458	HLKOKKSMTPARERENORMIVCNOMKYLLDKIDKEKOAKRKRRESVEOKRSRONMT 517
OY	2471	KLSALLFKHKEDLRRAETLKRRALLDKIDLQIEVDEELKRD 2510
DB	518	KLSALLFKHKEDLRRAETLKRRALLDKIDLQIEVDEELKRD 557
RESULT 4		
ABB68735		
ID	ABB68735	standard; Protein: 976 AA.
XX	ABB68735:	
AC		
DT	26-MAR-2002	(first entry)
XX		
DE	Drosophila melanogaster polypeptide seq ID NO 32997.	
XX		
KM	Drosophila; developmental biology; cell signalling; insecticide;	
XX	pharmaceutical.	
XX		
OS	Drosophila melanogaster.	
XX		

PN WO200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PMD, Myers EW;
 PI WPI: 2001-656860/75.
 DR N-PSDB; ABL12838.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 32997; 21bp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 976 AA;
 Query Match 10.9%; Score 1557.5; DB 22; Length 976;
 Best Local Similarity 36.4%; Pred. No. 1.3e-71;
 Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;
 QY 4 EEEEDGDAAEFOTDSEDEDEDEDDDDSDYPEMEDDDDDSYCTESSFRSHSTYSS 63
 DB 99 ESEHYGSDFGSDSESDSDNEDDMLTPPSDESLVANESSEPSVC--SFNONGV-- 152
 QY 64 TPGRRKRPVHRSPRIIEE-KDIPLEFPKSSEDLMPNHEIMNVIAIYEVNFGTVLR 122
 DB 153 ---GRPPRPSPRPVWLOEGROYAALDIPSSSEDLFIANTHVLAALSTYELRRFRHMVR 209
 QY 123 LSPRFEDFCALVSQEOCTMAEMHVLKAVLREBDTSNTTGPADLKDVSNTLYFI 182
 DB 210 LSPRFEDLCALACEQSALTEVHIMLKALIREBDAQCTHFGPDKDXTVNTISLYLI 269
 QY 183 DGMWPEVLYRYCESDKY----HHVLRYOEAEDYRPGVANKIKVLOFVDOGLTINIA 238
 DB 270 DSTWPEVLYRYSDSKTFEDRNPHILSHTE--YPTGIDNRLEVLOFLSDGLTISNI 326
 QY 239 REELMSEGVIOYDHCRCYCHLGLDLCETCSAYUHLCEVAPLEVEPEDEMOCEVCVAH 298
 DB 327 RDVMOEGRIHYDHCRCYCHLGLDLCETCPAYUHLCEVAPRPNVDVPTEDMOGGLCRSH 386
 QY 289 KVPGVTCVAELQKKPYIRHEPIGYDRSRKKYFLNRLIIEEDTENENKKTIWYSTK 358
 DB 387 KVGAVDCVLFOEKQGVLIIRHDSLGVDHGRKKYFIARIRPI-EDQENFT--CWYSETI 442
 QY 359 VOLAELIDLCIDKDYEAELCKLILEMREIHRHMDITEDLNRKSGSKSLAANETIL 418
 DB 443 SKLILLRLDAELETLSQITERRDELRQMKLTETLTNEHNRKRSYI----- 494
 QY 419 ESIRAKKGDIDNVASPEETEKDKNTENDSKDAEKNREPEFDOSEKSDDKTPDDDEQ 478
 DB 495 -----ETEGEAKN-----ELLEKEVLDDEDEKO----- 516
 QY 479 GKSEYGDGFKSEKSGELSESFGACKGASGSTRITIRLRNPDSKLSQLKSQVAAAHAHEAN 538

DB 517 -----GDAKSESQSIIE-----GTCKOEECKMVR-----QKSNQLT----- 547
 QY 539 KLEFKEKEVLVNVSCOELISRLSKKREVIKGNINNNFKILOGEKYRVYNOYSTNSFALN 598
 DB 548 -----NGLTFEKLGMEOGFRKNYNOYSTNFIALN 576
 QY 599 KQHREDHDKRRHLAHRFCLTPAGEFKWNGSVYKSLTISTRLTLITOLENNIPSSFLH 658
 DB 577 KPQNEERDKRRRLSHKFSILTASDFKIMGTDDNNITTLRQTLINFEESNIAASFNL 636
 QY 659 PNNASHRANWIKAVQMCSPREPALALILECAVYKVMPLPIREPLGTRILHRMSTIR 718
 DB 637 INNVVNRKKIWNNAVNMNARPSERFAVLLLFQASIKLSVFNANWHEOLGHTTLQRTISAER 696
 QY 719 EEEKYKKKKKKKKO---EEETMOQATWVYTFPVKHOVWKOGKEERYVYGGMWSIKRT 775
 DB 697 EERKKLEKREKRRRDEEBEENRLAFNYIKTTLGKIQVWKQKGEERYVIGQMCWMLSS 756
 QY 776 ----HYREFVPKLPQNTVNVYRKSLEGTKNNDENWDESDKRKRSRSPKKIKIEPDER 830
 DB 757 RRCGVARRARQPLTHNRVYVHYTM--GEENDVNEII-----LVDPRTQRFMOQCES 805
 QY 831 DEYKGSDAKGAQO--NENDISKITEKKODVKEILLDSDSDKCKREPEHVEDDMKTESH 888
 DB 806 SNVDGQVCHYLPDQYKRVKVIEDVTE-----KIKKH 836
 QY 889 VNQOESSQVNVVNSGFFHL--RTSYKKKTSSKIDGLLERRIKQPTLEKQRLKIKLE 946
 DB 837 -----IDYSKALNAPGRTIYSKVAKRKRIDDLDRLRKLALVEEQQ--ASKIP 882
 QY 947 GGIKIGICTSTNSSKNISES 966
 DB 883 SDMKPLIVSSQNTMTANSKQT 902
 RESULT 5
 AAB54364
 ID AAB54364 standard; Protein; 328 AA.
 XX
 AC AAB54364;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:816.
 XX
 KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.
 XX
 OS Homo sapiens.
 XX
 PN WO200055320-A1.
 PD 21-SEP-2000.
 PF 08-MAR-2000; 2000WO-US05989.
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-579444/54.
 DR N-PSDB; AAC99129.
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particularly pancreatic

XX BLanchard AD, Champion B, Page KR;
PI WPI: 2002-306794/35.
DR N-PSDB; ABR47549.
XX
PT New SKAT-2 polypeptide transcription factor, useful for treating or
PT preventing e.g. allergy and Alzheimer's disease, also related nucleic
XX acid, antibodies and modulators
XX
PS Disclosure: Page 37-38; 47pp; English.
XX
CC The invention describes an isolated SCAM/KRAB protein associated with a
CC Th2 phenotypic (SKAT-2) polypeptide (I). (I), especially where expressed
CC in recombinant cells, is used to identify agents that modulate SKAT-2
CC activity, and may be useful therapeutically. The agents, also (I) and
CC the polynucleotides (II) that encode it, are useful for treatment and
CC prevention of diseases responsive to SKAT-2 modulation, specifically
CC asthma; allergies (hay fever, atopic dermatitis; allergic rhinitis etc.);
CC Alzheimer's disease (and other neurological disorders, particularly where
CC related to apoptosis); conditions involving T-helper responses, and
CC apoptosis; also in vaccination against an antigen to enhance/repress Th2
CC or humoral responses, to reduce development of a Th1 phenotype, and to
CC manipulate the immune response in autoimmune diseases, infections and
CC cancer. (II) is also used for recombinant production of (I); as source
CC of antisense therapeutics and as primers and probes, e.g. for diagnostic
CC detection of mutations and for monitoring SKAT-2 expression in
CC association with disease. This is the amino acid sequence of the peptide
CC 7A1, that binds specifically to the kruppel-like zinc finger protein,
CC SKAT-2.
XX
SQ Sequence 149 AA:
XX
Query Match 5.5%; Score 789; DB 23; Length 149;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2633 GIIQSEAEILDEYVCPQOSTEDAMTVLPLETKDYEGKRLVRLSLOAHKMAFPLEPVD 2692
DB 1 GIIQSEAEILDEYVCPQOSTEDAMTVLPLETKDYEGKRLVRLSLOAHKMAFPLEPVD 60
QY 2693 PNDAPYGVYIKRPMPLATEERVORRYEKLFEVADMKIKIDNCRYYNPSDPFYQCA 2752
DB 61 PNDAPYGVYIKRPMPLATEERVORRYEKLFEVADMKIKIDNCRYYNPSDPFYQCA 120
QY 2753 EYLESEFVOKLKGFKASRSHNNKLOSTAS 2781
DB 121 EYLESEFVOKLKGFKASRSHNNKLOSTAS 149
XX
RESULT 9
AAU31944
ID AAU31944 standard; Protein: 246 AA.
XX
AC AAU31944;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2435.
XX
KW Human: vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; hematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US08656.
XX
PR 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
PI Tang YT, Liu C, Drmanac RT.
XX
DR WPI: 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 529-530; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate hematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukemias. AAU29510-AAU33304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 246 AA:
XX
Query Match 5.4%; Score 781; DB 22; Length 246;
Best Local Similarity 67.8%; Pred. No. 1.6e-32;
Matches 162; Conservative 10; Mismatches 63; Indels 4; Gaps 2;
QY 751 HQVWKQGEERYVYGVGWSMISKTHYRFPVPLPGNTNRYRKSLEGTNNNDENDES 810
DB 7 HQVWKQGEERYVYGVGWSMISKTHYRFPVPLHGNTNRYRKSLEGTNNNDENDES 66
QY 811 DKRCKSPKRIKIEPSEDEYKGSAAAGADONENDIKTEKKDQDKELLDSDSK 870
DB 67 DKRCKSPKRIKIEPSEDEYKGSAAAGADONENDIKTEKKDQDKELLDSDSK 126
QY 871 PCKEPEVDDDKTESHVNCQSSQVD--VVNVSEGFHLRTSYKKKTKSKLDGLLERR 928
DB 127 PCKERTN--GSRXHXERKVCKLGEFSSRCQCXXCFSSKDXLQKENKITOTRWISXKE 184
QY 929 IKQFTLEKQRLKIKLEGIGIKGTSTNSSKNLSSEPVITRAKECCQSDSKROEOP 987
DB 185 NXYVYVYGRKATATRRKNQYGMNXXGRKOSTNSSKNLSSEPVITRAKECCQSDSETRTER 243
XX
RESULT 10
ABB62887
ID ABB62887 standard; Protein: 997 AA.
XX
AC ABB62887;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide spQ ID NO 15453.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX

PT interactions -
XX Dislosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
PS The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB12072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2768 AA:
SO
Query Match 4.0%; Score 572.5; DB 22; Length 2768;
Best Local Similarity 18.4%; Pred. No. 2.7e-20;
Matches 562; Conservative 437; Mismatches 1147; Indels 905; Gaps 128;
QY 78 PLEEKDIP---PLEPKSEDLMPNEMHINVIATYEVLRNFGTVLRSPFEDFCA 134
DB 84 PITTCEPVVDVPHYSPEPTELSTPEK---GCSIEKFPYBGAOVSPNPKPCCLCYC 140
QY 135 LVSOECLTMAEMHVLLKAVLREEDTSTNTFGPADLKDSVNSTLYFIDGMTPEVLARY 194
DB 141 INNGTKVMQ-----BCTLH-VQG-----CLPIY 163
QY 195 GCSDEYHNVLYQEDYPTGPNVKNKIVLOFLVDQILTN----- 236
DB 164 NNGS-----CCCVRYSCDH-----ENELD-----FMDSTTTTIVRPTTGFLASTMP 209
QY 237 -----IAEELMSEGV-IQYDHCYVCHL-GDILC-----TCSAVYHL 275
DB 210 PTTTDCIHGELFPADASLCKGNACCHCYCMRGDLYCAVQCEVPMMAANGSCRMPPA 269
QY 276 ECVKPLEEVPDEMOCYEVCAHKGVTDCV-----AEIOKNKPIRHEPIGYRSRR 329
DB 270 BECCPSNVVCDSDSTTIVEITPESATSVPAKGIHAEIKEDVDLDDDDKNKE 329
QY 330 KTWFLNRRIITEEDTENENEKIKIWIYSTVQLAELIDCDKDTW---EAECLKILEMRE 386
DB 330 TATISSAEIGSGEIEEIEEEKD---KATVAVPVQVTD-EKDFSEFSESTAGIPSDSR- 383
QY 387 ELHRHMDITEDIANKARGSKSFLAANEELIESIRAK-----GDDDNK----- 432
DB 384 -----IDLPSSTEESEKSTEAEDIVKIVTTPPEGSGEEDVPRPSQIPEKIT 434
QY 433 -----SPEE-----TEKDKNET-----ENDSKD 450
DB 435 EDELIVKTSAPAKASPEEEVYATTSAPTEDVKTPTAGTISEEEBEKPTPAEGSGEE 494
QY 451 AEKN-----REEFEDQ-----SLEKSD-----DKTPDDPEQKSEV 483
DB 495 EEKDVKVTAPARETDEAKPTSAVPADEKEQEPKSESGDELDLKPPTAPATGATSA 554
QY 484 GPFKSEKSGELSESP-----CAGKAGSSTIILRLRNPDSKLQSKQOVA 531
DB 555 SESEBODGCKSTEAPTSVDITIEPAKPTESSSEASGEEDVAKETTPAGEASTAGEEIV 614
QY 533 AAHAHANKLFKGEKVLV-----VNSQGEISRLST-----KREVIK 568
DB 615 KCTTPAGESSSGDEEIVKGTTPAESSESEDELDLKTTPPAGEPVSAGEEELAKTTTA 674
QY 569 GINNYFKLQEGCKRYVHNQYSTNSFALNKQHREDHDKRRHLAKFCLTPAGEKMWG 628
DB 675 GPPS---IAGEEIEIVK-----TTPAGESSTIAGEEIVK-----VTTPAGESSESG 717
QY 629 SVHSGKVLV-----ISLRLIITOLENNIPSSFLHPNASHR 665
DB 718 EEEITKVTTPAGESSESGDEEIVKSTPAGEPIPSGEEDVIKATYSAPKSDIEGVKEPET 777

QY 666 ANWIK--VOMCKSPREFALALALECAVKPVYMLPIMREFLGHTRLHRTSIEREK 723
DB 778 ATEVPAEEVEDFAK-----TTPIAEEEPILAGTPIPDGIS-----GEEEIVK 822
QY 724 VKKEKKOEELTMOQATVKTTPPVKHQVWKOKGEERYTGTGCMWSIKTHVTFVK 783
DB 823 GTTPQLEQOPETSESTEV---PVAED-----D 848
QY 784 LQNTVWN-YRKSLEETKNMDEMDESDKRCRSRPPKIKIEPDEKD----- 831
DB 849 LSSSTISASAIASSTBQVQASSETTSSAPARAGDKDEAATVPTPADCKDEVEDATDL 908
QY 832 ---EVKSDAPADONEMDISKITEKDDVYKELDSDPKCKEPEVDDMK----- 884
DB 909 PVEDVQSTTAKTITTQCKEESSTEADEAEIEVTTSSPADQVEVAPADKDHDEED 968
QY 885 -----TESHVNCOESSQVDVNVVSGFHL-RTSYKKTKSKFLDGL 924
DB 969 VQTATDLPIKSDIGPVVDTEATTGQPTSDETATDKPPSVYLPVYQEVPSSTAKVDNR 1028
QY 925 LERRIKQFL-----EKKORLEKIKLEGIG-----IGKSTSSKNLSSPVITKA 972
DB 1029 NDETEKPTLPSPGSDQSSPLPAMDLPAGIPBEGDCLVBGKYANNVTVPATAPCDVSC 1088
QY 973 KEGCOSDSMKROESPANNDQPE-----DLIOGCSQS---DSSVLRMSDPSTHTKLYP 1023
DB 1089 K-CISLIVACQMECKLPENLEKCTVAADLDGCCPTIYICDESTSAKDEESTAK--P 1144
QY 1024 KDRVLDDVSTRPEKCPKONSINDIEK-----VSLASRGQEPYKSKTK 1070
DB 1145 DNKIDDVSEISTE-EIPDVIIMPGTITQPLSHVKNPDEIOPTVSVPAQFESTAKVD 1203
QY 1071 GDNFIDDSKLASADIGLICKNKKPLQIESDITVSSSKSALHSVSKSTDRBPPL 1130
DB 1204 KRP--IDES-----AEKRPIGESED-----SKPIDSEEDKKV 1237
QY 1131 SRAMDEKLGCDSESNSTLENSDPTVSJODSSEEDMIY-----QNSNESISEQFTR 1183
DB 1238 E-----ESAEKDKPVDESEKPKLPVTIPASEIEKSPDEDEKKT 1278
QY 1184 BOD---VEYLEPKCELYSGESTGNCEDRLPYKGTGANK---KPSQOKLEERPYNKC 1236
DB 1279 EADFAAPTEQPEATPPAQIADTAEKLEVDDKLATTSAPVSGEDELKPADKKRTE----- 1332
QY 1237 SDOI---KLKNTTDKKNENR-----ESEKKGQTSN--FQINKOKPKIYLGECLKEI 1287
DB 1333 TAOIPDAEIPASTDEPESSTELPTVLDKKPEDSYKGTGAPBSOKVPEVPTSASTENEI 1392
QY 1288 SESRVVSGHVEPKVNNINKIIP--ENDIKSLVKE-----SAIRPFIINGDI----- 1332
DB 1393 EESDKFTTAPAKRISASDETEPTAEEDLVPAFTEPILSEFEVSTKRPVNOGPPILPLA 1452
QY 1333 -MEDFERNSETKSHLLSSDAEGNYRDSLETLPSFKESDSTOTTP--SASCPESNVN 1390
DB 1453 OPEKKRPVDAETSTEADISTEPSAEVEKEASGETSESDNEIDAGASSTPVVPSADEKTPS 1512
QY 1391 QVEDMEIETSEVKVYSSIT--SEESN---LSNDIDENGLPJ-----KNENVNEE 1439
DB 1513 TEKTYE---ADCKFTTVAPLAGDEESNLPKLPQDIFEEBA-PVAVTTAPASDD--GE 1565
QY 1440 SK-----RKTVTEVTTMTSTVATESKYIKKEGDKQTVVSTENCAKSTYT 1487
DB 1566 OKPVEVEEKPRIEDGQKPIDEDTSTPS-----SENIEPESDRATIIASKEE--PSERS 1618
QY 1488 TTTTIVTKLSTPSTGGSVDIISVKEOSKTVVTTVTDLSLTGGTLVISMVYKEYSTD 1547
DB 1619 TGAPTKDEPAPESSTDAPESEDS--KETPESEVPTTAPA---GEKIPSTISITPDEPTAT 1673
QY 1548 KVKIMKFSRPPKTRSGTALPSYRKFTKSTKSIIVLPPNDLKLARKGCIREVFPFN 1607
DB 1674 SAPVAKPDEDVEKETSTELTPDAPASSSEEDNS---STDQIPS-----EVP---E 1717


```

QY 1608 AKPALDWIPSPRPFGITWRYRLQTVKSLAGVSLMLRLMALSRLRMDMAKAVPPG-- 1665
Db 1718 KKP-----LTPAOTPEBGDI 1732
QY 1666 -GSTRETSETETTTETIIRRDVGPYJFEXCIKRIICPGVPTPKETPPQK-GL 1723
Db 1733 VGATAAFTTSDVPPVQRLPE-----EVLAIPIQPSTETGTI 1768
QY 1724 RSSALRKRETPKQIGPVIIEIETWVAEEELMEIRAFARVER--EKAQAVFQ-----Q 1776
Db 1769 KQDETTAAISIDRK-----EPYVEIDEAETTVAPISKRDKPTEEEKPVQKPTGEE 1822
QY 1777 AKKRLKQOKPTVIATSTSTSTSTSPISPAOKVMVAPISGVTTGTAKMLTTKVGSPAT 1836
Db 1823 PSEEEKEKEPIEDVSEGVSTIASAGSTES-----SEFKPSTEGEVAEKPEB--- 1873
QY 1837 VTFQONKNFQHTAATWVKGOSNS-GVVOYQOKVGLIIPSTG--TSQOTFTSFQPRTA 1892
Db 1874 -----KQPSSTAQAAPVETIPEISTELPAQDDDKPTSEAPVDSDEDTSA 1916
QY 1893 TVTIR-FNTSGSGGTTNSQVITGPOAIRPGMTVIRTPLOOSTLGKAIIRTPVMWQCAPQ 1951
Db 1917 PSDEKIPSVSG-----EEVSGPEV-----TTASPOAAEBDELKTPASEPSTTD 1960
QY 1952 QV-WTQIIRQOPVSTAVSAPNTVSTTPGOKSLTSAVST-----SNTQSSASOPPRPOQG 2005
Db 1961 KVPETEYOKBEDETKADETESVATQVSDVATSTSAVPAAGDIEKDEQATASP--EIEEE 2018
QY 2006 VKLTMAQLTQITQGHGNGQLTVVITGOGOTGQLOLIPQGVYLP--GGGQOLMQ-AAM 2062
Db 2019 IKPTIAFAAEIPQ-----PSEKEPVDE-QEVEGTRKATPAESDQPIDELIAPA 2065
QY 2063 PNGVQRFLEPLATATATATATTTTSTAGTGEOGROSKLSPOMVHODK-----TL 2116
Db 2066 TSGIDE-----ASTAPRKESTIYASAA-----SP-AVHDEIKDVTYTG 2106
QY 2117 PPAOSSSVGPAKAQOPOTA-----QPSAR-PQOPQOPQSPAQPEVOTQPEVOT--QTV 2166
Db 2107 PVADEKEVAAPQODETKTISIDVSTDSPTAODEKODKTEAPVAPPTVSSPADSADSTP 2166
QY 2167 SSHAUS-----EAOPTH--AQSARKQVA-AOSOPQSNVQGSPPRV--QOSFORIRPS- 2215
Db 2167 TVEVPSPEVIDTKPMDIMSQTIAPHTADGAASTDEDEDAPVTVSPQDAEKTPVPAP 2226
QY 2216 -----TSQSLSPGOQSOQVOTTTSQ-----PIPIQPHSLQIPQSGOPQSPQVQSSSTQ 2264
Db 2227 QDSQKTSSEKAPQADEIPATAPFLDNKIPATVAPAPQTDGVPATAPFLDEDKIQTTAAP 2286
QY 2285 L-----SSGQTLNQSVSSSPRQLOIQOPQPOVIAVPOLOQOVQVLSQIQSOVVAQIQ 2318
Db 2287 LDEEKIPSTAPLDEKIPAPVSPVFPVPEPSEKPAVSEYDGE-----ESTEPVHDVE 2341
QY 2319 AQQSGVPOQITLOLPPIQIQGSSAVQTHQIQNVTVQAA-----SVQOLQGVQOOLRQOQ 2373
Db 2342 TSTBEPPTSDAKLPPTSAIPSPATPESPATETAIPEITAPPELEKEVPEKATEQPLEKETP 2401
QY 2374 KKKQOQIEIKREHTLOASNOSEI-----IQKQVYMKHNAVIEHLKOKKSNTP 2420
Db 2402 EKATEQPLEKETPEKATEQPLEKETPEKATEQPLEKETDQATQEPESVDEKTTPEP 2461
QY 2421 AER-----EENQRMIVCNQVMKYILDKIDKEEQAAKKRRKREESVEQKSKQNAIKISA 2474
Db 2462 VVKPSLIDSTERDE-----ESVESEESADKKKDKKKEETDETDKHEEPEVPA 2508
QY 2475 L-----LKKHKEQLRAELIKKRALLDKLOLQIEVQELKRDLIKKEK 2516
Db 2509 VVSEIIPQSEBAVPTTGHPLPH--LASSTTTPAVDDR-----VGBE-----DEN 2553
QY 2517 DLMQLAQATAAACPVPY-----PVLPAPAPPPSPPPVQOHTGLSTPTLLVVASQKR 2572
Db 2554 TTVKLSSTTTSTSTSPYTSAPSTTTVASQOQOPIPTPPYV--HA-----PEYDEY 2603
QY 2573 KREERKSSSKKKKMISTTSKETKDKTLYCICKTPTYDESKFYIGCDRCQNMVHGRCV 2632

```

```

Db 2604 DEEEVPGPCRYAGKLYVASAQIIPRDDEPCDFCF------FRSDII 2644
QY 2633 GIIQS-----EALIDBYCPCQO-----STEDAMTVLTP 2662
Db 2645 CIIQSCPPPIAGCHEEPIISGFCCPRYECPSMAAVALNTTSTTTSTILTP 2695

RESULT 12
ABB65772
ID ABB65772 standard; Protein; 5533 AA.
XX
AC ABB65772;
XX
DE 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 24108.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL09875.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5533 AA;
XX

Query Match 3.9%; Score 560; DB 22; Length 5533;
Best Local Similarity 19.3%; Pred. No.3.le-19;
Matches 520; Conservative 382; Mismatches 948; Indels 836; Gaps 118;

QY 325 DRGRKKYWFNIRRLIEDTENENKKIWIYSTKVQLAELIDCLDKDWEALCKTLEEM 384
Db 1997 DKEQKEKEIKREKDLREKQREKREK-----ELRDKDLREKEM-REKEOR 2041
QY 385 REETIRHMDITE-----DITNKARGSNKSFLAANDEILIESIRAKGIDIV 431
Db 2042 EKEIHKREKQREHREKREKQSRAMVDVEQGRGM-----RELSSYQKSKMDIAGE 2093
QY 432 KSPETEKQKNETEMDSKDAEKNEEFEDQSLK-KSDSDKTPDDDDPEQKSE----- 482

```


Dd	2094	AS	-----SLTIDQNHKKENAMOTIAQGTGASPTSPDNTREKSRKLSRNP	2142
Oy	483	VGDFEKSENGEISESESPACKGASGTR	-----ITRLRNPDSK-----LSQLKSQOV	530
Dd	2143	VLRLKRRLLSOSSEHNSHSGGCGSCGSHQIHNEHDYKRIIMENSONISVHNSNQRLNDR	2202	
Oy	531	AAAAHEANKLFKEBEKVLVYNSOCESR	-----LSYKK-----	563
Dd	2203	DSKEKSSS-FKEDK-----NSSSHISRPHCGSSASSSKHHNRDRHNOKGSASJET	2256	
Oy	564	-----EVI-----KGNINNYFKLQ-----BQYVYVYNQSTNSPLMK-----HOHR	603	
Dd	2237	NSSIEVYVDPISQTKHNLNTSEELOSHOPKREKEBEHNSHANSSSSKHKSRDHNNR	2316	
Oy	604	-----EDHCKRRHLAHFKCFLTPAGEFKMNSVHSGKVLITSLRLITPLE	649	
Dd	2317	EKKRHSVAESTNTDEHTPOQHNPFR-RLSAG-----SSSAG-----ELSSAA	2359	
Oy	650	NNIDPSFLPHNMASHRANWIKAVOMCSKPREPALALILECAVKPVMLPWREFLGHTR	709	
Dd	2360	TNTSGKLIHQ-----HHRSVRERKSSRGSDGHHSSKSL-----RAK	2398	
Oy	710	LHRMTSITEREK-----EKVKKKKKKOEELTMOQATVKTTF	747	
Dd	2399	LMNLSSASDQTDADKSKHSIFDIPDCPYNVSYDKVAKRSCKNMORAEKIKAKFS-	2457	
Oy	748	PKVHQVQKQKEEYRVYVYGMS-----WISKYHYRVPALPNTVYVNRKSLGKKNMD	804	
Dd	2458	-----QLQSRAKKRRKRSYSDGDDTEFEDROH-----RNSSSSFHGR--YPLGSSDD	2505	
Oy	805	ENMDESDKRCRSRPPKIKTIEPSEKDEYKGSDAKAGADONE-----MDIKSTEEKQDQ	859	
Dd	2506	DDDETHQRRIS-----SDSDAHEGQDGNQAGASTLDANVRQMQ-QN	2547	
Oy	860	VKELIDSDPOCKEERMEVDMDMKESHYNQOESSQVYVNVVSG-----FHLRSTY	912	
Dd	2548	LRLRCDGDDS-----SED--ELTRRVNMKHSHGKNSNSTRIASDSQSOQAPADLTIKQH	2602	
Oy	913	-----KKKTKSKSLDGLLERRIKQFTLEEKORLEKIKLEGKIGKIGKSTNSKULS	964	
Dd	2603	PIAPOELKRRQSLDEQOKRKSHDNSSTEER-----KAKTREL-----KTELDFYMS	2634	
Oy	965	ESPVITAKBEGCOSDSKROQSPNANDOPEDLIQCSQSDSS-----VLMSDPS--	1015	
Dd	2655	EYVTGKLEKSEPTERRKHKHKSRR-----LKSSSTADTSAQTPLVMTPLTPSIFD	2706	
Oy	1016	-HTTNKLIKPLDRVLVDVST-----BSPEKCKKONSIENTDIEKYSDLAS	1059	
Dd	2707	VHSSSEKTEKFDNDDIKTECSSIPLEISAGERKHKEREKREKRLKMTAEATVPN--	2763	
Oy	1060	RGOEPTKSKTGNDFIDDEKLSADDIGTLICKNKKPLIQEESPTIYSSSKSLHSSVP	1119	
Dd	2764	---SPTINDT-----SSEKLSKEERHRLKSKSKSKMDNSCNTKIYNSGA-H--P	2808	
Oy	1120	KSTNDRATPLSR-----AMDEGKLGCSSESNTLENSSDTJSIODSEEDM	1167	
Dd	2809	STSPSLPATPTSASTAQSTKREGDAMEFIPLIIODEESQCPQEAFLNKIDLPSS	2864	
Oy	1168	IYVONSNESISQFTPRQDVLEVLEPLKCLVSGESTGNCEDRLPYKGTGANGK-----KPS	1223	
Dd	2865	-VSTTGPVSAALOTYKQ-----EP-----STPNBS-----KNEBAIQLTVHEPE	2903	
Oy	1224	QOKLLEERPYNKSCDQIKLNTTDDKKNMNERSEK-----KOQRTSTPOJNKDKKPKXYL	1279	
Dd	2904	QOQOOLER---SRLSGGSSSSSHADRRHRREKREKREKRSQEDONQIHOKSSK----	2955	
Oy	1280	KGECLKEISBRYVSGNVEKRYVNNIKIIPENDIKSLYKESAIRPITNGVIMEDFENR	1339	
Dd	2956	-----VEKKVDDQNSVMDGAGRLEAQ-----LMSDPK	2966	
Oy	1340	NSSETKSHLSSDABGNTRDQSLLETLPSTK-----ESDSTQTTTPASCPSPNSVNOV----	1392	
Dd	2987	PISE-----EAPPTAANYRASMDVFEFFSON-IDNNSVMTKQGV	3026	

QY	1393	--EDMEIETSVYKVVYVSSPTTSEESNLSDPI	DENG	LPI	NNNEVNGESKRAKVI	IYEV	1450																										
Db	3027	KSEOECHKSKDKKKKKRSEKOEKLL-----	QOQR	RESL	PVVA	3067																											
QY	1451	TMSTVATTESTVYKVGKQKOT-----	VYSS	TENCAK	SVV-----	TTTTTV	1493																										
Db	3068	STSSAPPLPGVLVYNVAASKHADLOLDAKHIS	SPVYCK	SPSPL	PCLIGDDDDAL	HPK	3127																										
QY	1494	TKLSTPYGQSVDIISVKEOSKTV-----	VTVTV	TDLSL-----	TTGGLTVS	-----	1536																										
Db	3128	AKPTTPSSRGMDGLTFRERKPRLLSPKPTPI	ANSSTL	STJOS	SAETPVSSGIV	SSSALA	3187																										
QY	1537	-----	MTVSK	EXT	1545																												
Db	3188	TTPTSTAGVSAAPRGDNLSPTSACKKESF	IGPGDGL	DRIS	SAVOSIS	AEFNS	3247																										
QY	1546	-----	RDKYK	LKMF	FRPKKTS	SGAL	SYKFFYKTKK-----	1579																									
Db	3248	TSLLDNIADEKPIVAPSPRATKPRDLKES	KSRVYTI	SOETES	AVSALL	ESFGTS	TT	3307																									
QY	1580	-----	SIFL	PNDD	KLARK-----	GGIREV	PYFN	MAKPL	1612																								
Db	3308	DYSLDGDMSVNELETPLVLAEDEBEALAAK	ILET	AG	RAS	ILERP-----	EMEP	3353																									
QY	1613	DIMVYRSP-----	RPTFG	ITWRY	L-----	OTVYK	SLAGV	SLMLRL	MLMASL	KRMDMAK	VPBG	1665																					
Db	3364	EAPRDPEAEISEPEVLEVLDEBELKAVQSL	KNEHM-----	DIK	ADTP	QSE	3412																										
QY	1666	GSTRTESETEITTELKRRDVGPRGIF	EFIC	IKRIT	IC	IP	CV	RE	TP	KETPR	POOK	GLRS	1725																				
Db	3413	RDLIDIT-DTEENPDE-----	ADSSGP-----	SLKID	ETV	SSSS	PEKS	ISNN	3454																								
QY	1726	SALRPKRREPKRGVILITVWAEBELMEI	RAE	REKKA	QAV	EQAK	KRL	EQK	1785																								
Db	3455	S-----	PPREP	ANIDT-----	PVES	OPKLS	NEST	PRO	SVIT	KL	3489																						
QY	1786	P-----	TVAT	ST	PT	SS	TS	TI	SP	AK	VW	APIS	GSV-----	TTG	KRMV	LTK	VGSP	1835															
Db	3490	PFLDTPKTVAGRLPRSPVKLEP	TLISK	QD	LV	QV	GL	PL	RA	HS	HTG	SGIS	AN	SVIN	DL	3549																	
QY	1836	TVTPQOKNFHQFATVYKQSGNSGV	VOYOK	VULG	IS	TS	GT	SG	QFT	SE	OR	PLATV	1894																				
Db	3550	SNVJSSCSNTSASATGASASIS	IFSG	SP	AS	QAM-----	POAST	PKQ	GP	IT	PO	AKIT	QSL	3606																			
QY	1895	TIRNTSGSGTSTN-----	SGVIT	GPQ-----	PRG	MTY	IR	P	LO	Q	SL-----	GK	AL	IRT	1941																		
Db	3607	IMOPRTISIREPRTREHVAUPLVLS	PSHNR	PO	SG	TV	MG	IR	AP	SH	SP	SL	SP	GR	VA	SR	3666																
QY	1942	-----	PVW	OP	AR	PO	AV	MT	OL	IC	RO	ST	AV	SA	BN	TV	SP	RO	KLS	AT	ST	NI	1991										
Db	3667	LVGOLSPRGKRMVQR-SPOQ-----	QV	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	3722							
QY	1992	QSSA-----	SOP	RR	PO	GO	V	KL	MA	L	Q	L	Q	L	Q	L	Q	L	Q	L	Q	L	Q	L	Q	L	2048						
Db	3723	PTSKVSVYORBNQ-----	VP	OR	SP	K	VA	E	Q	T	PO	LM	3758																				
QY	2049	VLPBGQOLMOAMPKSTOVARE	LY	PL	AT	TA	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	2108						
Db	3759	I---P	L	K	M	T	R	I	O	V	R	N	H	R	T-----	L	I	S	K	U	V	U	R	O	A	S	A	PL	GR	3805			
QY	2109	QVNDKTLTPROSSSVYRAKAPOT	AD	SA	R	PO	R	T	OR	SO	R	AE	U	TO	R	E	U	TO	R	E	U	TO	R	E	U	TO	R	E	U	2167			
Db	3806	NVH-----	L	N	A	H	O	N-----	Q	O	O	R	U	L	A	K	I	T	A	N	O	O	N	O	O	N	O	O	N	O	3849		
QY	2168	SHVSEAPRTNAOSKFEUVALASOR	OS	PO	S	N	Y	O	G	S	R	U	V	A	K	O	S	R	U	V	A	K	O	S	R	U	V	A	K	05	2226		
Db	3850	-----	Q	O	O	L	H	G	S	O	Q-----	Q	I	S	A	R	O																

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PS Disclosure; SEQ ID NO 27225; 21bp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 6815 AA:

Query Match 3.8%; Score 549.5; DB 22; Length 6815;
Best Local Similarity 17.7%; Pred. No. 1.4e-18;
Matches 611; Conservative 346; Mismatches 1236; Indels 1063; Gaps 146;

QY 2 VSEEEEDDADAEETODEDEDEDE---MEEDDDSDYPEEMEDDDDDASYCTESSFRS 57
DB 1436 IDNOKTKKISEPPTSESDSTIEEYPRKISEHDEDLQTOEYSDVDSL-----P 1486
QY 58 HSTYSSTPGRKRRVYRPR-----PILEP-----KDIPLEPKSSEDLMP 100
DB 1487 KSKRRSKKQKESLPISLYTRLEETPEPTTEKTYEDGKVVYVINKRIRKKRAGP 1546
QY 101 NEHNMATYAE-----VLNRFQVLRSLSPFRFD-----FCAA 134
DB 1547 KEYIEVETIETEDNNPADVYIRIETTPSIDSKPRQEDHKIQYQEKPTESIDNYIQ 1606
QY 135 LVSGE-----QCTLMAE-----MHVLLKAVLRSDTS 162
DB 1607 LIDDEIQVHKEPKATVLETSPESKAKKIKKHKKTTEVLDIPIVLEVIQIETED 1666
QY 163 NTTFGRPD--LKD-----SYNSTLYFDGQTPREVLYAVGSESKYHNVL 205
DB 1667 DEDKPRPEVLTKEIDHENAEAPKEQVEITEKTAIDEPISVAVTEITDQEPQ--EVL 1724
QY 206 PYQEAEDYPRGYENK-----IKVLOFLVD--QFLTNTIAREELMSGVIOYD 251
DB 1725 PAQGEKK-----PIKQKKLKPEDVNTVYVLELETEPQF-----ETIPEDA---D 1768
QY 252 DHCYVCHKGLDLCCEFCSAVYHLECYKPRLEEVPEDEMOECVANKVGVTDCAVEIQ 311
DB 1769 DK-----POVIEDISENVQVOLIIEEDGTPK---QVEIK 1800
QY 312 KNRKYIHPIGYDRSRKRYW-----LNRLLIIEEDEN- 346
DB 1801 KKVYSPKHGP-----KEQVFETTERPSDEPLAEVTYELTEEGJLKDVIPIQEKTV 1853
QY 347 ENKKIYVSTKYQALAILDCLDYWAELCK--ILEMBREBH-----RHMDI 394
DB 1854 KKPRLKLPEDIQSVIVVLEEFNEPQWPASTEKRIIDIASIEIIVPTEEDGITKEVEV 1913
QY 395 TEDLTNKAQSGNK-----SFLAANEELLESIRA 423
DB 1914 KKKVYSRKQGTKNQVFEIETKTSDEPLAEVTYIELSGDSQSEVTLIPKREKPIKTKIKL 1973
QY 424 KKGIDID-----NKSPEETEKDNKTENDSKDADAKNREFEDQSLEKSD 469
DB 1974 KPDDVESTVAVVLEEFCEPOSFESPEPTGPAHTTKTKKPKPKPIYKAPENWIL--IEE 2031
QY 470 KTPDDDEQGSKEVG--DFKSEKNGELSESPGAGSGSTRITTRLNPDLSKLSQLSQ 528
DB 2032 MAPTVLENIYVNEIGEYKQVTKTKLKKEG-----PREYILEIKE- 2073
QY 529 QVAAAABHANK-----LFEKG-----KEYLVVNSQGEISRLSTKKKVKMGK 570

DB 2074 -----TYENKREGDIEITTELVEGSPDASDDPVIYVQ-----KIKKKKPVKD 2120
QY 571 INNYFK--LGQEGKYVYVNOYSTNFSALNKHQHREDHDKRRHLAKKFLTPAGEKMG 628
DB 2121 LDKYIOQLIEQ-----IKTPLEYEPTLMDSKK-----PKKVK--- 2157
QY 629 SVHGSKVL--TISTLRLLITOLENIPSSFLHPNMAASHRANWIKAVOMCSKPREFALAI 687
DB 2158 -SHNKKTIVIGLPLTY-----HEEN---VEDIVSEPDIMETPKTL 2195
QY 688 LECAYKPVVMLP-----IWRP-LGHTRLHMTSIEREKEKVKKKKKQKQ- 732
DB 2196 LD-EIKEIPQLPDDSSKYLVINISDEGEADKIPQTOQPIKKEKPLKKKDVPEVSL 2254
QY 733 -----EETMOQATWVKTTFPVKHQVWKKQEEYRVYGGGMSWISTHYRFPV 782
DB 2255 EAFDHTVAVYSEPTLDEGT--VKEVYKRRKVRKRG-----SDHIFELT- 2297
QY 783 KLPGNINVYRKSLEGT--KNNDENMDESDK---RKCSRPKKIKIPEPSE----- 829
DB 2298 ---ETTSDBRPLAEVTVVLELSDVLDSEKPKHERKLVKKPKQLKDKDVEYIINILE 2353
QY 830 -----KDEYKGSAAKADQONEMDISKI--TEKKDQ-----VKELLDSDSKP 871
DB 2354 EFTQPIPVGLVEDEYERKVKQETKPKKSPITYIATEQEDNNNNYDALKEDLDPIERA 2413
QY 872 CKREPMVEDDDMKTESHVNOQSSQVDVYVNSGFLRTSYKKKTSSKLDLLEIRI-- 929
DB 2414 LEKPSPLETYSVEDSSGEEQKOPKPKKISKPKIKQPSVDKSDYLIVNYSSEIID 2473
QY 930 ---KQFTL---EKKQ-----RLERIKLEGKIGKSTNSSKNLSSEPVITAK-- 973
DB 2474 EPIPEYVYVTEAEKRPSEPTPKVELETE---AVEKEVDDDDGGETTKOSVTRKIK 2529
QY 974 -----EGCOSDSMRQEOFPANNDQ-----PEDLIQ 999
DB 2530 KLVGPREELIETVETGDTPEYEVITTEEYQEKKEPEEKAKTVAKKAKIPADQ 2589
QY 1000 GCSQSDSVLRMSDPHTTKLYPKRDVLVDVYSRSPETKCPKONS-----IEND 1049
DB 2590 DYTQK---LIBODIKTELEKYEKIDDEPVKMKRKPPIKYYQSGEQKETEPEIDEK 2645
QY 1050 IEKVSDDLASQGE-----PYKSKTKGDNFFIDSKLAS---ADIDIGL---I 1091
DB 2646 PVKEISEYVSDDEPKLTVAVKEFIPEKPEEPFIVYLEEVESKRPDEGKVRKY 2705
QY 1092 CKNKKPL-----IOESDITVSSSKALSHSVKSTNDRATPLSR----- 1132
DB 2706 VKTKIKQNRGSEVYVHDIVLEIDTDPNRESVITVTTVPLETPDQOPSVKQKRTKIK 2765
QY 1133 ---AMDEFGKLGCDSESNSTLENSDVTYSIOD---SSE--EDMIVONSNEISQFRT 1183
DB 2766 DEVEDFVYKRV--IEEAPQEGSVLDLVIEDFVFKSSSEKRRKKPKIKDKHTSVEE--TPH 2822
QY 1184 EOVLEVLE-----PLKCELVGSESTGNCEDRLPVYGTEAN-----GKPSQOKK--- 1227
DB 2823 EDEVLILIESVPEDSPISDDLIT-----VDSVPILIEEPENKYNQIEDTKPKPKKPRP 2876
QY 1228 -----LEERPVKSCDOJK---LKNTTDKNNENRESEK 1258
DB 2877 SAKILEBNPDETVKPLALHTDSDEKPDVOEFSISIKEEQKITHHEKKSSKISSE 2936
QY 1259 KQORTY--FOINGKDNKPKIYLKGCLEKISRSRVVS--GNVEPKVNNINKIIPENDIKS 1315
DB 2937 QRPQOSTQEOYEIVTEHD---LKPEEKRFYQVYQISEFNVEETKDDQGVY-----HKQ 2987
QY 1316 LTVKESAIRPEFINGDVIMEDFNERNSSEFKSHLSSSDAEGNRSDELFLPSTKESDSIQ 1375
DB 2988 VTKRMLRMRAGAGEIIEIVVRDDQPEAEITIVVEPEPVRV--QDEKPKPEPKK---TR 3042
QY 1376 TTTPSASCPSNSVNOYQEMEITSYKKTSSPISSEESNLSNFIDENGPIPKNEN 1435


```

Db 3043 KYKKDDIHVIQKLIIELEPKTELEKIEIEPEPIYKOKPLDSPIDVLDSEKPEVOKKDK 3102
Oy 1436 VNGESK---RKVTIEEVTMTSTVATESK-----TVIKVEKGDQVAVSTENCA 1482
Db 3103 KSKSTVPEHEETPVQEQYAKVNVVEEAEQPEIPVQIIEVKEVEADVEVI---TED-G 3158
Oy 1483 KSVVTTTTTIVKLSPSGSGSVDIISVEQOSTVTTTIVTDSLTGTLVTSMVSKSE 1542
Db 3159 KPVQEKTRRVKLKIGREBQTEFKITMIESENDSVTVIIVDEE-----PEIASPOSIEE 3212
Oy 1543 YSTRDVKLMKESRPKTKSTGALPSYRKFWRKSTKKSIFVLJPNDDLKLKLRGRIEVP 1602
Db 3213 HEQOSEKLA--PKPKTYRKAKKDDLSYVKKLIEEL---PKVLEKY-EKVEPEKXP 3266
Oy 1603 YFNYNAPRALDIMPYSPRPRTGCI---TWRYRLQTVKSLAGVSLMLRLMLASLRMDMA 1658
Db 3267 -VKLVSDSIPREPPKPKDSQPSISVLBDTTRPKKTKTPK-----TPKIEDND 3311
Oy 1659 AKVPPGGSTRTESE-TEITTTETIKRDVCGYGR-----PEYCIKRIICP 1705
Db 3312 QOVPDEPTETVDTNDIPRLTPTQTAQPDGTATQITPSAQEEKSTQDDTKDTIQKTV-- 3369
Oy 1706 IGVPEPKEPTPQKGLSSALRPKREPQKQTPGVIIEETWVAEEL--ELMEIRAFAP 1763
Db 3370 -----KHKTKPDQKSVETSEL-----PEVHKDYQISIIHEELVEEPEKLIIEVVIDE 3420
Oy 1764 RVEKEKAQAEQOAKKRLDQKPTVATSTSTSTSTSTISPAQ-----KVAV----- 1812
Db 3421 VAEVESQPIVEEVEDEEPOPAETEVEEDVTPKSKSKKKVKKKTDHDELKMKMEQEI 3480
Oy 1813 -----APISGSVTGKMLTTLTVGSGPATVTFQONKN 1844
Db 3481 EKTELEKYKIEFDVPKKLREPALEPIKIEREOKPKTKVLTADTVKPKVKKLPSKR 3540
Oy 1845 FQOTFATWKGQSGNSGVQVOQKVLGIIPSSGTSGQFTSPQPTATVTTIRPNTSGSG 1904
Db 3541 -----KEKPAEELTVQ-----LPKRLKARMLVEYRP---APLIPKTTIDIG 3579
Oy 1905 GTTNSQVITG-----POLRGMVTIATPLQOSTLCK-----AII RTPV 1943
Db 3580 AIKDNELSRNIEAEELIKFPHKTKIKIKIDLEKVELEKYEKVISSEEPKEKTPY 3639
Oy 1944 MVOFGAP-----QOVMTOIIRGQPVSTAVSANVYSSITPGOKSLTSANSTSIQSSAQ 1997
Db 3640 KRPKRPKPEEKOEDVKLIGKAKKPKKEEAPENV-----TLKNI----- 3680
Oy 1998 PRPOOGGYKLTMAOLTQTOGHGNOGLVVIQGGQGTGQLOLPI-----QGVTVLP 2051
Db 3681 POKPQVEEVEELKQKPKVE-----IVEQTKPKDGEVVEPEPSEFDRPEYVP 3732
Oy 2052 GPGQOLMOAMPNGTVQRFPLTATTAATTTTSTTAASTGEORQSKLSPQMOYH 2111
Db 3733 DELEQIEHEIPE-KVKKPSKTKYKPKDKSKSEPEITVEIYAGVPKKEEA--IPQDQW 3789
Oy 2112 QDKTLPPAOSSSVGPAKKAPQTAQPSARPQOTQOPSPAQPEVQTOPEVQTOPTVSSHVP 2171
Db 3790 FRKPERDAEEDNDSEIKLRP-VPOASKDENPDEQALVTPKAEPRIPQELIEDAIDDEKRP 3848
Oy 2172 SEAPOTHAOSSKPOVAAQOSQPNVQO-----SPVAVQSSQRIKRPSPQSLSPOQ 2224
Db 3849 KSKPKKQVOPKQDELAKKEPEEVEVSKKEEALVDPKPIETEKPKDVKVEKKPKKE--AP 3905
Oy 2225 QSOVOTTTISQ-----IPIQPH--TSLQIPSGQPOPOPOV-----QSTOTLSSG 2268
Db 3906 VSEVYVIEEPEPEVEPIEYKITTIVLEBEDAKPHQKAYVIDFDERQETTEVEIEE 3965
Oy 2269 QTLNOVSVSSPSRP---QIQIQQPOPOVAVPOLQOQOVVLISQIOSQVVAQIQAOQSGVP 2325
Db 3966 KVVTRKKKPKPOQPEFEFETLKEPKKEE---QIQPQV-----VSAEISLPIEPEQO-KP 4014
Oy 2326 QOIKLOLPI-----QIQQSSAVQTHQIQNV-----TVYQAASVQEOLOLR-VQOL 2368
Db 4015 EOYEVELKITTOTTPPEPNDVQIAVKEKVTKPKVKKEDKIVVVEAEKEKPEVETIVEV 4074

```

```

Oy 2369 RQOQORRKO-----QOIEIK-----REHTIQ 2389
Db 4075 EKQEEKKSEKPKSYEFKISQISIEKPLEVAEAPPEETPKVVEKVAEKFSYSFTLK 4134
Oy 2390 ASNQSEIT-----OKOVYMKHNAVIEH-----LKQKKS 2417
Db 4135 ETDEKRVITVDQPEBEAPVEVFKKRPKEPAVEAEFWTEPKVIEETSVETAIKQKKT 4194
Oy 2418 MPFAREENORMI-----VCNOVM-----KYLKIDIOKEEK-----QAKKR 2454
Db 4195 KPKKDEEERQALAIKVESEAPVAEEVSEAPSKVIEEVEVIAEERPKETTRVSESEK 4254
Oy 2455 KREESVEQ---KRSKONATKLS--ALLFKHQBQLRAEL-----KKRALD---K 2496
Db 4255 PEESPVEQFTVKKRKSVPFADPRATEIYIKESKRAVEVTEADHAIKTKPKKKVTQVEAE 4314
Oy 2497 DLQIEVOEELKRDLIKKEKDMQLOAATAVAAP-----CPPTVPLVA----- 2540
Db 4315 ELKIKITEEVPOEIPILIEEVEEVEVITETKTAAPVVEEKTYKIGIKETEPKPAEAVEE 4374
Oy 2541 --PPAPP--SPPPRGVQH--TGLLSP-----TLPVASOKKREBEKD----- 2579
Db 4375 EEPVYTEPIEAPKRPVEECHKVRYVLETPRELVEEVEKAVIRKKRPRIKEPEPA 4434
Oy 2580 --SSSKSKKKMISTTS-----KETKDKTKLYCI-CKTPYDESKFYIGCDRCQMW 2626
Db 4435 EYVTSRPKVEEVEATSSIAVIEQPTBEEADKLTIIIEETPPQE----- 4481
Oy 2627 YHGRCVGIQ---SEALDIEYVCPQOCSTEDAMTVTLPTLEK--DYEGIKRVILRSIQANK 2682
Db 4482 -----LVQZIEIEIIEVEPKAPBEPDND--FTFATKQSEKRPVEELPBEQVTIQKK 4532
Oy 2683 MAMPFLEVPDNDAPDYGVGIEKPMDLATMEERVOR 2718
Db 4533 KRAPVEVVEEBPA-EFLVKPKTPVQDVTBEAKITK 4567

RESULT 15
ABB64877
ID ABB64877 standard; Protein; 3080 AA.
XX
AC ABB64877;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 21423.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PA 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PMD, Myers EW;
XX
DR WPI; 2001-65660/75.
XX
DR N-PSDB; ABL08980.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

```


THIS PAGE BLANK (USPTO)

Qy	61	YSSIPGKRKRVRHPRSPILLEKODIPLEPRKSSSDLAMPNPHIMNVIAITEVLRNCTV	120
Dp	61	YSSIPGKRKRVRHPRSPILLEKODIPLEPRKSSSDLAMPNPHIMNVIAITEVLRNCTV	120
Qy	121	LRLSPFEEDFCALVSOECCTLMAEMHVLKAVLREEDTSNTTFGPADLKDSVNSTLY	180
Dp	121	LRLSPFEEDFCALVSOECCTLMAEMHVLKAVLREEDTSNTTFGPADLKDSVNSTLY	180
Qy	181	FIDGMTPEVLYRVYCESDKRYTHHVLRYQBAEDYRYGVENKIKVLQFVLVDFLTNNARE	240
Dp	181	FIDGMTPEVLYRVYCESDKRYTHHVLRYQBAEDYRYGVENKIKVLQFVLVDFLTNNARE	240
Qy	241	ELMSESVIYQDHDHCRCVHKIGDLLCCETCSAYVHLECYKPRLEEVPRBDMQCYCVAUKY	300
Dp	241	ELMSESVIYQDHDHCRCVHKIGDLLCCETCSAYVHLECYKPRLEEVPRBDMQCYCVAUKY	300
Qy	301	PGVYDCAEIOXKKPYIRHEPIGYDSRRKRYFVLNRLIIEEDTENENCKIKWYSTVQ	360
Dp	301	PGVYDCAEIOXKKPYIRHEPIGYDSRRKRYFVLNRLIIEEDTENENCKIKWYSTVQ	360
Qy	361	LAELIDCLDKDYBAELCKILEEMREEIHRHMDITEDLTNKAAGSNKSFLLAANEILLES	420
Dp	361	LAELIDCLDKDYBAELCKILEEMREEIHRHMDITEDLTNKAAGSNKSFLLAANEILLES	420
Qy	421	IRAKKGIDVWKSEPEEKXKNTENDSDAKENRPEFDOSLEKDSODKTPDDOPEQG	480
Dp	421	IRAKKGIDVWKSEPEEKXKNTENDSDAKENRPEFDOSLEKDSODKTPDDOPEQG	480
Qy	481	SEVDFSEKSNGLSESPAGKAGSGSTRILYRLNPNPSKLSQKLSOOVAAAHEAKL	540
Dp	481	SEVDFSEKSNGLSESPAGKAGSGSTRILYRLNPNPSKLSQKLSOOVAAAHEAKL	540
Qy	541	FKEGEVLYVNSOGELSRLSSTKKEVIMKGINNYFKLGOEGKYRYVHNOYSTNSFALNKH	600
Dp	541	FKEGEVLYVNSOGELSRLSSTKKEVIMKGINNYFKLGOEGKYRYVHNOYSTNSFALNKH	600
Qy	601	QHRDHDHKKRLHLAKFCLTAPGAFKNGSVSHGSKVLTITLTLTTLTOLENNIPSSFHLPN	660
Dp	601	QHRDHDHKKRLHLAKFCLTAPGAFKNGSVSHGSKVLTITLTLTTLTOLENNIPSSFHLPN	660
Qy	661	WASHRAWIKAVOMCSKPREFALALILBECAPVYMLRIMREPIGHTFLHMTSIEREE	720
Dp	661	WASHRAWIKAVOMCSKPREFALALILBECAPVYMLRIMREPIGHTFLHMTSIEREE	720
Qy	721	KEVKKKEKKOEBEETMOATWVYTFPVYKHQYWKOGGEYRVYGVGWSWISKTHYRF	780
Dp	721	KEVKKKEKKOEBEETMOATWVYTFPVYKHQYWKOGGEYRVYGVGWSWISKTHYRF	780
Qy	781	VPLKPGMNTNMYRKSLEGTNNMNDENNDSDKKCRSKRIKIPBDSKDXCVKSSDAK	840
Dp	781	VPLKPGMNTNMYRKSLEGTNNMNDENNDSDKKCRSKRIKIPBDSKDXCVKSSDAK	840
Qy	841	GADONEMDISKITEKQOADVKELDSDSPCKKEEPMEDDDMKTESHNCQESSQOVV	900
Dp	841	GADONEMDISKITEKQOADVKELDSDSPCKKEEPMEDDDMKTESHNCQESSQOVV	900
Qy	901	NVSEGFHLRTSYKKKTKSSKLDIGLLERRIKOFTLEBKORLEKIKLEGGIKIGIKTSTSS	960
Dp	901	NVSEGFHLRTSYKKKTKSSKLDIGLLERRIKOFTLEBKORLEKIKLEGGIKIGIKTSTSS	960
Qy	961	KNLSESPVITAKAGCCSDMRQDOSPANNANDPELDIGCSQSDSVYLRMBDSHHTNK	1020
Dp	961	KNLSESPVITAKAGCCSDMRQDOSPANNANDPELDIGCSQSDSVYLRMBDSHHTNK	1020
Qy	1021	LYPDRVLDVYSISPTKCPKONSTIENDIEEVVSLAARGOEPKSKIKGNDFTIDSK	1080
Dp	1021	LYPDRVLDVYSISPTKCPKONSTIENDIEEVVSLAARGOEPKSKIKGNDFTIDSK	1080
Qy	1081	LASADDIGTLCKNKKRPLIOEESDTIVSSSKASLHBSVSKSTNDRDAPFLSRAMFBEKL	1140
Dp	1081	LASADDIGTLCKNKKRPLIOEESDTIVSSSKASLHBSVSKSTNDRDAPFLSRAMFBEKL	1140
Qy	1141	GCDSBSNSTLENSSDIVSIQDSSSEEDMIVONSNEISIEQFRTREODVEVLEPKCELVS	1200
Dp	1141	GCDSBSNSTLENSSDIVSIQDSSSEEDMIVONSNEISIEQFRTREODVEVLEPKCELVS	1200

Dd	1141	GDSDSNTLENSDPTVSIQDSSSEEDMALVQNSNESISQFPTRODVEDLEPKCELVSQ	1200
Qy	1201	ESTGCEBRLPVKGTGEANGKRPQOKLEBRPVNKCSDQIKLNTTDKNNENRESEKQ	1260
Dd	1201	ESTGNCEDRLPVGKTEANGKKRPQOKLEBRPVNKCSDQIKLNTTDKNNENRESEKQ	1260
Qy	1261	QRTSTFOJNGKDNKPKITLKGCBLKEISESRVSGNBPKNVNNINKIIPENDISLTYKE	1320
Dd	1261	QRTSTFOJNGKDNKPKITLKGCBLKEISESRVSGNBPKNVNNINKIIPENDISLTYKE	1320
Qy	1321	SAIRPFINGDIWMEPFENRNSSETKSHLSSSDAEGNRDSELETLPSTKESDSQOTTTPS	1380
Dd	1321	SAIRPFINGDIWMEPFENRNSSETKSHLSSSDAEGNRDSELETLPSTKESDSQOTTTPS	1380
Qy	1381	ASCPESNSVNOVEDMEIETSEVKRVTSPTISEESNLNDPIDEKGILPKNNENVGES	1440
Dd	1381	ASCPESNSVNOVEDMEIETSEVKRVTSPTISEESNLNDPIDEKGILPKNNENVGES	1440
Qy	1441	KRKIVITEYTTMTSVATAEKSTVIVKEGDMQVYVSTSENAKSTVYTTTTVYKLSPS	1500
Dd	1441	KRKIVITEYTTMTSVATAEKSTVIVKEGDMQVYVSTSENAKSTVYTTTTVYKLSPS	1500
Qy	1501	TGGSVDIISVKEOSKTVVTTVTDSLTJTGGLVYTSMTVSKESYSTRODKVLMKESRPKKT	1560
Dd	1501	TGGSVDIISVKEOSKTVVTTVTDSLTJTGGLVYTSMTVSKESYSTRODKVLMKESRPKKT	1560
Qy	1561	RSGTALPSYRKRFVTKSTKKSIFVLPNDDLKILARRGGIREVYENYNAKPALDIWPPSP	1620
Dd	1561	RSGTALPSYRKRFVTKSTKKSIFVLPNDDLKILARRGGIREVYENYNAKPALDIWPPSP	1620
Qy	1621	RPTBGITWRKYRLQTYKSLAGVSLMRLMLASLRMDMDMAKXVPPGGSGRTSTSETEITTT	1680
Dd	1621	RPTBGITWRKYRLQTYKSLAGVSLMRLMLASLRMDMDMAKXVPPGGSGRTSTSETEITTT	1680
Qy	1681	EITKRDDGPGVGIREFYCAKRTICEIGVETPEKPTPOQRGLRSASLRPRKPEPKQTG	1740
Dd	1681	EITKRDDGPGVGIREFYCAKRTICEIGVETPEKPTPOQRGLRSASLRPRKPEPKQTG	1740
Qy	1741	PVLIETWVAEEBELMEIRAFERVEKEKAQAEQOAKKRLBQOKPVIATSTSPISST	1800
Dd	1741	PVLIETWVAEEBELMEIRAFERVEKEKAQAEQOAKKRLBQOKPVIATSTSPISST	1800
Qy	1801	TSTISPAOKVMWAPISGSYTTGTKNVLTTTKGSPATVTFQONKNHOFEFATWVWOGGONS	1860
Dd	1801	TSTISPAOKVMWAPISGSYTTGTKNVLTTTKGSPATVTFQONKNHOFEFATWVWOGGONS	1860
Qy	1861	GVVOYQOKVLGIIIPSSGTSGOOTFSPQPRATVIRNTSSGGTINSQVITGPQIRP	1920
Dd	1861	GVVOYQOKVLGIIIPSSGTSGOOTFSPQPRATVIRNTSSGGTINSQVITGPQIRP	1920
Qy	1921	GMTVIRTPLOOSTLICKAIIIRPVMVOPGAPQOVMQIIRGQPVSTAVSAPMTVSSTPGOK	1980
Dd	1921	GMTVIRTPLOOSTLICKAIIIRPVMVOPGAPQOVMQIIRGQPVSTAVSAPMTVSSTPGOK	1980
Qy	1981	SITSATSTNSIOSSASQPPRPQOGYKLTMAJOLJOLGHHGNSGLJYVIOGGQTTQOL	2040
Dd	1981	SITSATSTNSIOSSASQPPRPQOGYKLTMAJOLJOLGHHGNSGLJYVIOGGQTTQOL	2040
Qy	2041	QLIPQGVNVLPPGGOLWQAAAMPNCTVORLETPPLATTATTASTTTTVSTTAACTGBOR	2100
Dd	2041	QLIPQGVNVLPPGGOLWQAAAMPNCTVORLETPPLATTATTASTTTTVSTTAACTGBOR	2100
Qy	2101	QSKLSPOKQVHODKTLRPAOSSSVPAKAPQPTAOPASARPQOTQPOSPAQPEVOGTQPEV	2160
Dd	2101	QSKLSPOKQVHODKTLRPAOSSSVPAKAPQPTAOPASARPQOTQPOSPAQPEVOGTQPEV	2160
Qy	2161	QOTTVSSHVPSAOPHTAOSKSPQVAAOSQPOSNVGQSPVNVQSPQOTRIRBSTPSQL	2220
Dd	2161	QOTTVSSHVPSAOPHTAOSKSPQVAAOSQPOSNVGQSPVNVQSPQOTRIRBSTPSQL	2220
Qy	2221	SFGQOSQVQTTTSQPIPIQPHISLQIPQOGQPOSQPOVQSSQTLSSQOTLNOVSVSPS	2280
Dd	2221	SFGQOSQVQTTTSQPIPIQPHISLQIPQOGQPOSQPOVQSSQTLSSQOTLNOVSVSPS	2280

Db 2221 SFGQSQVOTTTSSQPIPIOPHTSLQIPSGQPOSDPOVOSTQTUSSLGGTILNOVSVSFS 2280
Qy 2281 RPOLOIOPOPOVIAVAPOLQOQOVVLSQIQSQVVAIQAOQSGVPOQIKLOLPQIQOQSS 2340
Db 2281 RPOLOIOPOPOVIAVAPOLQOQOVVLSQIQSQVVAIQAOQSGVPOQIKLOLPQIQOQSS 2340
Qy 2341 AVQTHQIQNVTVQAASVQEQLOQVQOQLDQOQKQOQIEIKREHTLQASNOSEITIQK 2400
Db 2341 AVQTHQIQNVTVQAASVQEQLOQVQOQLDQOQKQOQIEIKREHTLQASNOSEITIQK 2400
Qy 2401 VVMKHNVAVEHLKOKKSMTPAREENQBMIVCNQVKKYILDKIDKEKQAARKKREESV 2460
Db 2401 VVMKHNVAVEHLKOKKSMTPAREENQBMIVCNQVKKYILDKIDKEKQAARKKREESV 2460
Qy 2461 EOKRSKQNTKLSALLFKHEDQLRAEILKRALLDKLOIEVQEBLRLKIKKEKDLQ 2520
Db 2461 EOKRSKQNTKLSALLFKHEDQLRAEILKRALLDKLOIEVQEBLRLKIKKEKDLQ 2520
Qy 2521 LAQATAVAAPCPPTVPVLPAPAPPPSPPPPGVQHTGLSTPTLPVASOKRREBKDS 2580
Db 2521 LAQATAVAAPCPPTVPVLPAPAPPPSPPPPGVQHTGLSTPTLPVASOKRREBKDS 2580
Qy 2581 SSKSKKKKMIITTSKTKETKDTLYCCKTPYDESKFYICGDCQNNYHRCVGILOSEAE 2640
Db 2581 SSKSKKKKMIITTSKTKETKDTLYCCKTPYDESKFYICGDCQNNYHRCVGILOSEAE 2640
Qy 2641 LIDEVYCPOCOSTEDAMVLTPLTEKDYEGKLVLSLOAHKAMPFLPPVDPNDAPDY 2700
Db 2641 LIDEVYCPOCOSTEDAMVLTPLTEKDYEGKLVLSLOAHKAMPFLPPVDPNDAPDY 2700
Qy 2701 GVKEPMDLATMEERVORRYEKLTEFVADMTKIEDNCRYNPSDFYQCAEVLESFV 2760
Db 2701 GVKEPMDLATMEERVORRYEKLTEFVADMTKIEDNCRYNPSDFYQCAEVLESFV 2760
Qy 2761 OKLGFKASRSHNNKLOSTAS 2781
Db 2761 OKLGFKASRSHNNKLOSTAS 2781

RESULT 2
US-09-698-295-1
Sequence 1, Application US/09698295
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
FILE REFERENCE: 06501-068001
CURRENT APPLICATION NUMBER: US/09/698,295
CURRENT FILING DATE: 2000-10-27
PRIOR APPLICATION NUMBER: PCT/JP99/02340
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: JAPAN 10/137631
PRIOR FILING DATE: 1998-04-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2907
TYPE: PRT
ORGANISM: Homo sapiens
US-09-698-295-1

Query Match 98.9%; Score 14170; DB 20; Length 2907;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2766; Conservative 2; Mismatches 13; Indels 126; Gaps 1;
Qy 1 MYSEEEEDDGAETQOSEDEDEDEDEDDDDSDYPREMEDDDDDASCTESSFSHSHT 60
Db 1 MYSEEEEDDGAETQOSEDEDEDEDEDDDDSDYPREMEDDDDDASCTESSFSHSHT 60
Qy 61 YSSTGRKKPRVHRSPRIIEEKDIPLEFPKSSSEDLAMPNEHINNVIAIYELRNFGTV 120
Db 61 YSSTGRKKPRVHRSPRIIEEKDIPLEFPKSSSEDLAMPNEHINNVIAIYELRNFGTV 120
Qy 121 LRLSPREFDFCALVSQOCTLMAEMHVVLLKAVLREEDTSNTTGPADLKDSVNSTLY 180

Db 121 LRLSPREFDFCALVSQOCTLMAEMHVVLLKAVLREEDTSNTTGPADLKDSVNSTLY 180
Qy 181 FIDGMTPPEVLRYCSDREYHNVLPYOEAEDYPYGVENKIKVLQELVDQFLTITIAHE 240
Db 181 FIDGMTPPEVLRYCSDREYHNVLPYOEAEDYPYGVENKIKVLQELVDQFLTITIAHE 240
Qy 241 ELMSEGVIOYDHCRCVCHLGLDLCCECSAVVHLEBCVPRPLEVEDEDMQCEVCAHV 300
Db 241 ELMSEGVIOYDHCRCVCHLGLDLCCECSAVVHLEBCVPRPLEVEDEDMQCEVCAHV 300
Qy 301 PGVTDCAVEIOKNKPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWYSTKVQ 360
Db 301 PGVTDCAVEIOKNKPYIRHEPIGYDRSRKRYWFLNRLIIEEDTENENKKIWYSTKVQ 360
Qy 361 LAELIDCLDKDWEALCILEEMREIHRHMDITEDLTKAKGSNKSFLAANEELIS 420
Db 361 LAELIDCLDKDWEALCILEEMREIHRHMDITEDLTKAKGSNKSFLAANEELIS 420
Qy 421 IRAKKGIDINVKSPETEEDKNETENDSKDAEKNREFEQSLKSDDKTPDDPEQK 480
Db 421 IRAKKGIDINVKSPETEEDKNETENDSKDAEKNREFEQSLKSDDKTPDDPEQK 480
Qy 481 SE----- 482
Db 481 SEPTREVGDKGNSVANLGDNTTNTATSEETSPSEGRSPVGLSETPDSSNMAEKVASEL 540
Qy 483 ----- 482
Db 541 PODVPEPNKTCSSNTSATTTSIQPLENSSSSSELSNQSSESAAADDPENGERSHT 600
Qy 483 -----VGDFSEKSNSELSSESPAGKAGSGSTRILRLRNPDSKLSQKSOQVAAA 534
Db 601 PVSIOEIVGDFTSKSTGBELSSPGAGKAGSGSTRILRLRNPDSKLSQKSOQVAAA 660
Qy 535 HEANKLFKSGKEVYLVNSQGEISRLSTKKEVIMKGINNYFKLQGEKRYVYHNOYSTNS 594
Db 661 HEANKLFKSGKEVYLVNSQGEISRLSTKKEVIMKGINNYFKLQGEKRYVYHNOYSTNS 720
Qy 595 FALNKHQHEHDHKKRHLNHPCLTPAGEFKMGSVHSGSVLITSLRLITTOLENNIS 654
Db 721 FALNKHQHEHDHKKRHLNHPCLTPAGEFKMGSVHSGSVLITSLRLITTOLENNIS 780
Qy 655 SFLLPMMASHRAWIKAYOMCSKREFALALILECAVPRVWVLPIMREFLGHTRLRHMT 714
Db 781 SFLLPMMASHRAWIKAYOMCSKREFALALILECAVPRVWVLPIMREFLGHTRLRHMT 840
Qy 715 STEREEKYVKKKKQOEETWQATWVKYTFPVKHQVWKQGEERYVYTGQGSWISK 774
Db 841 STEREEKYVKKKKQOEETWQATWVKYTFPVKHQVWKQGEERYVYTGQGSWISK 900
Qy 775 THYVRFVPRKLPQNTNVNRYKSLBGTKNMNDENDESDKRCSPKIKITIEPSEKDEVK 834
Db 901 THYVRFVPRKLPQNTNVNRYKSLBGTKNMNDENDESDKRCSPKIKITIEPSEKDEVK 960
Qy 835 GSDAAGADQNEMDISKITEKKDQVYKELLDSQDKPCKEPEHREYVDDDKTBESHVACQS 894
Db 961 GSDAAGADQNEMDISKITEKKDQVYKELLDSQDKPCKEPEHREYVDDDKTBESHVACQS 1020
Qy 895 SOYDVVNVSEGFHLRTSYKKTSSKLDGLERRIKQFTLEEKQKLEKIKLEGGIKGK 954
Db 1021 SOYDVVNVSEGFHLRTSYKKTSSKLDGLERRIKQFTLEEKQKLEKIKLEGGIKGK 1080
Qy 955 TSTNSSKNLSESPVITKAKEGQSDMRQDSQSPNANNQDPEDLIQCCSDSDSVLPMSP 1014
Db 1081 TSTNSSKNLSESPVITKAKEGQSDMRQDSQSPNANNQDPEDLIQCCSDSDSVLPMSP 1140
Qy 1015 SHTTNKLVPKRVLDVDSIRSPETKCPKONSIEENDIEEKVSDLASGQPTSKTGKGF 1074
Db 1141 SHTTNKLVPKRVLDVDSIRSPETKCPKONSIEENDIEEKVSDLASGQPTSKTGKGF 1200
Qy 1075 FTIDSKLASADDTIGTLICKNKKPLIQEESDTIVSSKSAHSSVPKSTNDRATPLSRAM 1134


```

Db 1201 FIDSKLASADIGTILCKNKKPLIOESDPTIVSSSKSALHSSVSPKSTNDRDATPLSRAM 1260
QY 1135 DEFGKIGCSESNSTLNSSDTVSIODSSBEDMTVONSNSTSEORTRRODEVLEPLK 1194
Db 1261 DEFGKIGCSESNSTLNSSDTVSIODSSBEDMTVONSNSTSEORTRRODEVLEPLK 1320
QY 1195 CEVJSESTNCNCDRLPVKQTEFANGKRRPSQOKLEERPVKCSDOJIKLKNTTUKNNENR 1254
Db 1321 CEVJSESTNCNCDRLPVKQTEFANGKRRPSQOKLEERPVKCSDOJIKLKNTTUKNNENR 1380
QY 1255 ESEKKGORSTFOINGKONKPKIYLKGECLKEISESRVSVGNVPEPVNNINKIIPENDIK 1314
Db 1381 ESEKKGORSTFOINGKONKPKIYLKGECLKEISESRVSVGNVPEPVNNINKIIPENDIK 1440
QY 1315 SLVKSARITPFIINGVIMDEFERNNSSETKSHLSSSDAEGNRPISLETPSTKESDPT 1374
Db 1441 SLVKSARITPFIINGVIMDEFERNNSSETKSHLSSSDAEGNRPISLETPSTKESDPT 1500
QY 1375 QTTTPSACPESNVNOVEDMEIETSEVKKVTSSPTISEESNLNDPIDENGLPIKNKE 1434
Db 1501 QTTTPSACPESNVNOVEDMEIETSEVKKVTSSPTISEESNLNDPIDENGLPIKNKE 1560
QY 1435 NVNGESKRKIVITEVTTMTSTVATESEKTVIKVKGDKQTVVSTENCASKTVTITTTVT 1494
Db 1561 NVNGESKRKIVITEVTTMTSTVATESEKTVIKVKGDKQTVVSTENCASKTVTITTTVT 1620
QY 1495 KLSPTSGSVSDIISYKESKTVVTTVTDLSITTTGTLVTSMTVSEKSTYTRDKVKLMKF 1554
Db 1621 KLSPTSGSVSDIISYKESKTVVTTVTDLSITTTGTLVTSMTVSEKSTYTRDKVKLMKF 1680
QY 1555 SREKRTSGTALPSYKRFVTKSTKSIIFVLPNDDLKKLARKGGIREVYFENYNAKPALDI 1614
Db 1681 SREKRTSGTALPSYKRFVTKSTKSIIFVLPNDDLKKLARKGGIREVYFENYNAKPALDI 1740
QY 1615 WPPSPRPRTGCTIMRWYRLQVVKSLAGVSLMLRLMLASLRDDDMAAKVPCPGSGSTRPETS 1674
Db 1741 WPPSPRPRTGCTIMRWYRLQVVKSLAGVSLMLRLMLASLRDDDMAAKVPCPGSGSTRPETS 1800
QY 1675 TEITTTTIIKRDVGPYIGFEEYCIRKIIICPIGVETPKETPTPORKGLSSALRPKPE 1734
Db 1801 TEITTTTIIKRDVGPYIGFEEYCIRKIIICPIGVETPKETPTPORKGLSSALRPKPE 1860
QY 1735 TPKQTGVIIEFWVAEELELMEIRAFAEVREKEKAQVBOQAKKRLBOQKPVVIATSTT 1794
Db 1861 TPKQTGVIIEFWVAEELELMEIRAFAEVREKEKAQVBOQAKKRLBOQKPVVIATSTT 1920
QY 1795 SPTSSTSTISPAOKVWVAIISGSVTGTGMVLTTKVGSPATVTFQONKNFHOTFAWVK 1854
Db 1921 SPTSSTSTISPAOKVWVAIISGSVTGTGMVLTTKVGSPATVTFQONKNFHOTFAWVK 1980
QY 1855 QGOSNSGVVOQKVLGIIISSTGTSOQFTSFQPRATVYIRPNTSGSGGTTSNSQVIT 1914
Db 1981 QGOSNSGVVOQKVLGIIISSTGTSOQFTSFQPRATVYIRPNTSGSGGTTSNSQVIT 2040
QY 1915 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVWVQCAPQOVMTOIIRGQPVSTAVSAPNTVS 1974
Db 2041 GPOIRPGMTVIRPPLQOSTLGKAIIRTPVWVQCAPQOVMTOIIRGQPVSTAVSAPNTVS 2100
QY 1975 STPOKSLTATSTNSIIOSSASQPPRPQOGVKTMAQLTOLTOGHGNGGLTVVIOGOG 2034
Db 2101 STPOKSLTATSTNSIIOSSASQPPRPQOGVKTMAQLTOLTOGHGNGGLTVVIOGOG 2160
QY 2035 QTTGOLILPOGVTVLPGPCOQLMOAAMPNGTVORFLFTPLATTATASTTTTSTTAA 2094
Db 2161 QTTGOLILPOGVTVLPGPCOQLMOAAMPNGTVORFLFTPLATTATASTTTTSTTAA 2220
QY 2095 GTGBOROSKLSPOMVHODKTLPPAQSSSVGPAKAQOPQAPQAPQOPQOPQAPQAPQ 2154
Db 2221 GTGBOROSKLSPOMVHODKTLPPAQSSSVGPAKAQOPQAPQAPQOPQOPQAPQAPQ 2280
QY 2155 QTOPEVOTOTTVSSHVSEAKOPHTHAOSSKPOVAAQSQPOSNVQSQSVRVQSPQOTRIRP 2214
Db 2281 QTOPEVOTOTTVSSHVSEAKOPHTHAOSSKPOVAAQSQPOSNVQSQSVRVQSPQOTRIRP 2340

```

```

QY 2215 STPSQLSPGOOSQVQTTTSQPIPIQPHSTLQIPSGQPOQPOQVQSSQTLSSGQTLNOV 2274
Db 2341 STPSQLSPGOOSQVQTTTSQPIPIQPHSTLQIPSGQPOQPOQVQSSQTLSSGQTLNOV 2400
QY 2275 SVSSPSRPOLOIQOPPOVIYANPOLQOQVOVLSQIOSOVVAQIOAOQSGVPOQIKLOLPI 2334
Db 2401 SVSSPSRPOLOIQOPPOVIYANPOLQOQVOVLSQIOSOVVAQIOAOQSGVPOQIKLOLPI 2460
QY 2335 QIOOSSAVQTHOLOVNVYVQOAAVBOLORVQOLRPOQOKKKQOQOIEIREHTLOASNOS 2394
Db 2461 QIOOSSAVQTHOLOVNVYVQOAAVBOLORVQOLRPOQOKKKQOQOIEIREHTLOASNOS 2520
QY 2395 EIIOQOVYMKHNAVIEHLKOKKSMTPAEREENQRMIVCQVMKYILDKIDKEEQQAARR 2454
Db 2521 EIIOQOVYMKHNAVIEHLKOKKSMTPAEREENQRMIVCQVMKYILDKIDKEEQQAARR 2580
QY 2455 KREESVEOKRSKONATKLSALLFKHKEQURAEILKKRALLDKLOLEVOEELKRDILKIK 2514
Db 2581 KREESVEOKRSKONATKLSALLFKHKEQURAEILKKRALLDKLOLEVOEELKRDILKIK 2640
QY 2515 EKDLMOLOAQTAAVAAACPPVTPVLPAPAPPPPPPPGVQHTGLSTPTLPVYASOKRRR 2574
Db 2641 EKDLMOLOAQTAAVAAACPPVTPVLPAPAPPPPPPPGVQHTGLSTPTLPVYASOKRRR 2700
QY 2575 EEEKDSSSKSKKKKMIITTSKETKDKTKLYCICKTPYDESKFYIGCDRCQNMWYHGRVGI 2634
Db 2701 EEEKDSSSKSKKKKMIITTSKETKDKTKLYCICKTPYDESKFYIGCDRCQNMWYHGRVGI 2760
QY 2635 LOSBAELIDEXVCPQOSTDANTVLTPLTEKYEBGLKRVLRSLQAHKAMAPLEFVDNP 2694
Db 2761 LOSBAELIDEXVCPQOSTDANTVLTPLTEKYEBGLKRVLRSLQAHKAMAPLEFVDNP 2820
QY 2695 DADPYGVVKEPMDLATEERVOORYEKLTEFVADMTKIFEDRCRYVNPDSFPYOCABV 2754
Db 2821 DADPYGVVKEPMDLATEERVOORYEKLTEFVADMTKIFEDRCRYVNPDSFPYOCABV 2880
QY 2755 LESFVQKLGFKRASRSHNNKLOSTAS 2781
Db 2881 LESFVQKLGFKRASRSHNNKLOSTAS 2907

RESULT 3
PCT-US02-2119-15
; Sequence 15, Application PC/TUS0221179
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: CANDHI, Ameena R.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YUE, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: GRIPPIN, Jennifer A.
; APPLICANT: YANG, Junming
; APPLICANT: SANJANMALA, Madhu M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: BOROMSKY, Mark L.
; APPLICANT: YAO, Monique G.
; APPLICANT: WALIA, Natinder K.
; APPLICANT: BANDMAN, Olga
; APPLICANT: IAL, PREETI G.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LIO, Wen

```



```
APPLICANT: TANG, Y. Tom
APPLICANT: ZEBARJADIAN, Yeganeh
APPLICANT: LU, Yan
TITLE OF INVENTION: Nucleic Acid-Associated Proteins
FILE REFERENCE: PF-1031 PCT
CURRENT APPLICATION NUMBER: PCT/US02/21179
CURRENT FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: US 60/300,518
PRIOR FILING DATE: 2001-06-22
PRIOR APPLICATION NUMBER: US 60/301,787
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,792
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,892
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/301,893
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/303,405
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/303,442
PRIOR FILING DATE: 2001-07-06
PRIOR APPLICATION NUMBER: US 60/364,438
PRIOR FILING DATE: 2002-03-12
NUMBER OF SEQ ID NOS: 72
SOFTWARE: PERL Program
SEQ ID NO 15
LENGTH: 2759
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 394992CD1
PCT-US02-21179-15
```

Query Match 93.7%; Score 13428; DB 1; Length 2759;

Best Local Similarity 90.5%; Pred. No. 0;

Matches 2635; Conservative 6; Mismatches 7; Indels 264; Gaps 3;

```
QY 1 MYSEEEEDDGAETODEDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFSHSHT 60
DB 1 MYSEEEEDDGAETODEDEDEDEDEDDSDYPEEMEDDDDDASYCTESSFSHSHT 60
QY 61 YSSTPGRRKPRVHRPSPILLEEKDIPLEPPKSSSEDLMPVNEHINVAIYEVLRNFGTV 120
DB 61 YSSTPGRRKPRVHRPSPILLEEKDIPLEPPKSSSEDLMPVNEHINVAIYEVLRNFGTV 120
QY 121 LRLSPREFDFCAALYSOQCTLMAEMHYLLKAVLREEDTSMTTGPRADLKDSVNSTLY 180
DB 121 LRLSPREFDFCAALYSOQCTLMAEMHYLLKAVLREEDTSMTTGPRADLKDSVNSTLY 180
QY 181 FTDMGWPEVLRVYCSDEKYEHNHVLRYQEAEDYRUGPVENKIKVLOFLVDQFLTNIARE 240
DB 181 FTDMGWPEVLRVYCSDEKYEHNHVLRYQEAEDYRUGPVENKIKVLOFLVDQFLTNIARE 240
QY 241 ELMSSEVIOYDHCVRCHKLGDLCCETCSAVYHNECVKRPLEEVRDEMOCEVYAHKY 300
DB 241 ELMSSEVIOYDHCVRCHKLGDLCCETCSAVYHNECVKRPLEEVRDEMOCEVYAHKY 300
QY 301 PGVTDCAVETLQNKPRYIRREPIGYDSRRKRYWFLNRLLIIEEDTENENKIKIYUYSTKY 360
DB 301 PGVTDCAVETLQNKPRYIRREPIGYDSRRKRYWFLNRLLIIEEDTENENKIKIYUYSTKY 360
QY 361 LAELIDCLDKDWEALCKILEEMREIHRHMDITEDLTNKAQSKSFLLAANEETLES 420
DB 361 LAELIDCLDKDWEALCKILEEMREIHRHMDITEDLTNKAQSKSFLLAANEETLES 420
QY 421 IIAKGGIDINVKSPEETEKOKNETENDSKDAEKNREEFDQSLKSDSDKTPDDPEQCK 480
DB 421 IIAKGGIDINVKSPEETEKOKNETENDSKDAEKNREEFDQSLKSDSDKTPDDPEQCK 480
QY 481 SE----- 482
DB 481 SEPTTEVGDKGNSVSNLQDNFTTNAETSEETSPSEGRSPVGCLETDPDSSNMAKKVASEL 540
```

```
QY 483 -----VDFEKSXNGELSESPGAGSGSTRILRLRNPDSKLQLSQVAAAAHEA 537
DB 541 PDVVPYGDFFSEKXNGELSESPGAGSGSTRILRLRNPDSKLQLSQVAAAAHEA 600
QY 538 NKLFGKEKVELVYNSOGELISRLSTKKEVIMKGINNYFKLGQKRRVYHNOYSTSPAL 597
DB 601 NKLFGKEKVELVYNSOGELISRLSTKKEVIMKGINNYFKLGQKRRVYHNOYSTSPAL 660
QY 598 NKQHREDDHKKRHLAHKFLPAGEFKMGSVHGSKVLTISRLTLTQLENNITSSFL 657
DB 661 NKQHREDDHKKRHLAHKFLPAGEFKMGSVHGSKVLTISRLTLTQLENNITSSFL 720
QY 658 HPNMAHRAHWIKAYOMCSKPREFALALILECAVPYVWLPIWREFLGHTRLHRMTSIE 717
DB 721 HPNMAHRAHWIKAYOMCSKPREFALALILECAVPYVWLPIWREFLGHTRLHRMTSIE 780
QY 718 REBEKRYKKKEKKOEDEETMOATWYKYTFPVKHQYWKQGEERYVYTGGSWISKTHY 777
DB 781 REBEKRYKKKEKKOEDEETMOATWYKYTFPVKHQYWKQGEERYVYTGGSWISKTHY 840
QY 778 YRFVPLPGNTNPNYKRSLEGTKNNDENMDESDKRSRPAKIKTEPDSEKDEYKGS 837
DB 841 YRFVPLPGNTNPNYKRSLEGTKNNDENMDESDKRSRPAKIKTEPDSEKDEYKGS 900
QY 838 AAKGADQNEMDISKITEKKQDYKELLSDSDKPKCEEPMEYVDDMKTESHVNCQESSQY 897
DB 901 AAKGADQNEMDISKITEKKQDYKELLSDSDKPKCEEPMEYVDDMKTESHVNCQESSQY 960
QY 898 DVNVSEGFHLRTSYKKRTKSSKLDGLLEBKIKOFTLEEKORLEKTLBGGIGIKTST 957
DB 961 DVNVSEGFHLRTSYKKRTKSSKLDGLLEBKIKOFTLEEKORLEKTLBGGIGIKTST 1020
QY 958 NSSKNLSESPVITKAKEGCOSDSMRQOSPANNDDPEDLIQGCSDDSVLMSPSHT 1017
DB 1021 NSSKNLSESPVITKAKEGCOSDSMRQOSPANNDDPEDLIQGCSDDSVLMSPSHT 1080
QY 1018 TNKLKPRYLDVSTRSPETKCPKONSIENTDIEEVSDIASLARGOEPKSKTGDNFFID 1077
DB 1081 TNKLKPRYLDVSTRSPETKCPKONSIENTDIEEVSDIASLARGOEPKSKTGDNFFID 1140
QY 1078 DSKLASADIGTLICKNKKPLIOESDITVSSSKSLHSSVPKSTYNDROATPLSRAMDE 1137
DB 1141 DSKLASADIGTLICKNKKPLIOESDITVSSSKSLHSSVPKSTYNDROATPLSRAMDE 1200
QY 1138 GRLGCDSESNSTLENSDPYISIDSSSEPMIYONSNESTSEQRTREPOVEVLEPKCEL 1197
DB 1201 GRLGCDSESNSTLENSDPYISIDSSSEPMIYONSNESTSEQRTREPOVEVLEPKCEL 1260
QY 1198 VSGESTGNCEDRLPVKGTTEANGKPKPSQOKKLEERPVNKCSDQIKLANTDKNNENRES 1257
DB 1261 VSGESTGNCEDRLPVKGTTEANGKPKPSQOKKLEERPVNKCSDQIKLANTDKNNENRES 1320
QY 1258 KKGQRTSTPOINGKDKPKIYLLKBECLKLIESRNVYSGNVEPVNINIKIIPENDIKSLT 1317
DB 1321 KKGQRTSTPOINGKDKPKIYLLKBECLKLIESRNVYSGNVEPVNINIKIIPENDIKSLT 1380
QY 1318 VKESAIRPFLNGDVIIMEDFERNSEFKSHLSSDAEGVRSLSLTLPSTKSDSTQTT 1377
DB 1381 VKESAIRPFLNGDVIIMEDFERNSEFKSHLSSDAEGVRSLSLTLPSTKSDSTQTT 1440
QY 1378 TPSASCPSNSVNOVDMETIETSEVKVYTSPTSEESNLNDPFDENGLPINKNENNA 1437
DB 1441 TPSASCPSNSVNOVDMETIETSEVKVYTSPTSEESNLNDPFDENGLPINKNENNA 1500
QY 1438 GESKRTVITEVYTMSTVATESKVIYKVEGDKQTVVSTENCAKSVYTTTTVTKLS 1497
DB 1501 GESKRTVITEVYTMSTVATESKVIYKVEGDKQTVVSTENCAKSVYTTTTVTKLS 1560
QY 1498 TPSTGGSVDIISKQSKVIVTTVTDSLTFTTGTLVTMTVSKKESSTDVKYLMKFSRP 1557
DB 1561 TPSTGGSVDIISKQSKVIVTTVTDSLTFTTGTLVTMTVSKKESSTDVKYLMKFSRP 1620
```



```
QY 1558 KTRRSALSYKRFVTKSRKKSIFVLPNDLKLARKGIREVPIYNNYAKPALDITWY 1617
Db 1621 KTRRSALSYKRFVTKSRKKSIFVLPNDLKLARKGIREVPIYNNYAKPALDITWY 1680
QY 1618 PSPRPFGITWRRLQYVKSILAGVSLMLRLMLASLRWDDMAAKPPEGSGSRPTSETEI 1677
Db 1681 PSPRPFGITWRRLQYVKSILAGVSLMLRLMLASLRWDDMAAKPPEGSGSRPTSETEI 1740
QY 1678 TTEETIKRDVGPYGINSEXCIRKIIICPIGVETPKETPTPORKGLSSALRPKRPETPK 1737
Db 1741 TTEETIKRDVGPYGINSEXCIRKIIICPIGVETPKETPTPORKGLSSALRPKRPETPK 1800
QY 1738 QTGPIVILETWAAEEELMEIRAFAEVEKEKAQVPOQAKKRLGQKPPVIATSTTSPT 1797
Db 1801 QTGPIVILETWAAEEELMEIRAFAEVEKEKAQVPOQAKKRLGQKPPVIATSTTSPT 1860
QY 1798 SSTSTISPAOKWAVIISGVTTGTGMVLITTKGSPATVFOONKMFHOTFATWVWQO 1857
Db 1861 SSTSTISPAOKWAVIISGVTTGTGMVLITTKGSPATVFOONKMFHOTFATWVWQO 1920
QY 1858 SNSGVVVOQKVLGIIPSTGTSGQFTSFQPRATVITIRPNTSGSGTSSNSQVITGPQ 1917
Db 1921 SNSGVVVOQKVLGIIPSTGTSGQFTSFQPRATVITIRPNTSGSGTSSNSQVITGPQ 1980
QY 1918 IRPGMTVIRPPLQOSTLGKAIIRTPVWVOPAPQOQVMTQIIRGQPVSTAVSAPVTSSTP 1977
Db 1981 IRPGMTVIRPPLQOSTLGKAIIRTPVWVOPAPQOQVMTQIIRGQPVSTAVSAPVTSSTP 2040
QY 1978 GOKSLTATSTNSIOSSASOPPRPOGOVKLTMAQLTOLQHGNGGLVIVIOGQOQTT 2037
Db 2041 GOKSLTATSTNSIOSSASOPPRPOGOVKLTMAQLTOLQHGNGGLVIVIOGQOQTT 2100
QY 2038 GOLQLIPQGVTVLPBGQOLMQAMAPNGTVORLEFPLATTAATTAATTTTSTTAAGTG 2097
Db 2101 GOLQLIPQGVTVLPBGQOLMQAMAPNGTVORLEFPLATTAATTAATTTTSTTAAGTG 2160
QY 2098 EOROSKISPMQYVHODKTLPPAOSSSGVPAKAPQTAQPSARPOQTOPSPAPQEVQTO 2157
Db 2161 EOROSKISPMQYVHODKTLPPAOSSSGVPAEAPQTAQPSARPOQTOPSPAPQEVQTO 2220
QY 2158 PEVOTOTVSSHVPESEKOPRHAOSKPOVAASOPOSNVGOSPVVRQOSQPRIRSTP 2217
Db 2221 PEVOTOTVSSHVPESEKOPRHAOSKPOVAASOPOSNVGOSPVVRQOSQPRIRSTP 2280
QY 2218 SOLSPQSOQOVQTTTSPQIPRIOPHTSLQIPISQOPQOSQPOVOSSTQTLSSQTLNQSVS 2277
Db 2281 SOLSPQSOQOVQTTTSPQIPRIOPHTSLQIPISQOPQOSQPOVOSSTQTLSSQTLNQSVS 2340
QY 2278 SPSPRQOLQIQPQPVYAVVQLOQOQOVVLSQIOSOVVAQIOAQOQGVPOQIKLOLPIQIQ 2337
Db 2340 SPSPRQOLQIQPQPVYAVVQLOQOQOVVLSQIOSOVVAQIOAQOQGVPOQIKLOLPIQIQ 2400
QY 2338 QSSAVQTHQIQNVVTVQASVQBOQLOVQOLRQOQOKKQOQOIEIKREHTLQASNOSEII 2397
Db 2400 QSSAVQTHQIQNVVTVQASVQBOQLOVQOLRQOQOKKQOQOIEIKREHTLQASNOSEII 2460
QY 2398 QOVVNMKNHNAVIEHLKOKKSMTPEAREENORMIVCNOVMKYIIDKIDKEEKOAKKKRE 2457
Db 2460 QOVVNMKNHNAVIEHLKOKKSMTPEAREENORMIVCNOVMKYIIDKIDKEEKOAKKKRE 2520
QY 2458 ESVEOKRSKONATKLSALLFKHKEQLRAELTKKRALLDKLOLEVOBELKRDLEKKEKO 2517
Db 2518 ESVEOKRSKONATKLSALLFKHKEQLRAELTKKRALLDKLOLEVOBELKRDLEKKEKO 2580
QY 2518 LMOLOATAAVABCPPTVLPAPAPAPPPPPPGVQHGLSTPTLPAASOKRKEE 2577
Db 2580 LMOLOATAAVABCPPTVLPAPAPAPPPPPPGVQHGLSTPTLPAASOKRKEE 2640
QY 2438 LMOLOATAAVABCPPTVLPAPAPAPPPPPPGVQHGLSTPTLPAASOKRKEE 2497
Db 2500 LMOLOATAAVABCPPTVLPAPAPAPPPPPPGVQHGLSTPTLPAASOKRKEE 2560
QY 2578 KDSSSSKKKKMTISTSKETKDKTLKYLCKTQYDESK----- 2615
Db 2640 KDSSSSKKKKMTISTSKETKDKTLKYLCKTQYDESK----- 2700
QY 2498 KDSSSSKKKKMTISTSKETKDKTLKYLCKTQYDESKFTIGDCLCNWYHGEVGTTEK 2557
Db 2560 KDSSSSKKKKMTISTSKETKDKTLKYLCKTQYDESKFTIGDCLCNWYHGEVGTTEK 2620
QY 2616 -----FTIGDCRCQNMWYHGRCVGIQOSEA 2639
Db 2680 -----FTIGDCRCQNMWYHGRCVGIQOSEA 2740

Db 2558 EAKKMDVYICNDCKRAGESSSEELYCIRPIYDESOPIYICDRCQNMWYHGRCVGIQOSEA 2617
QY 2640 ELIDEEYVCOQOSTEDAMTVLTPLTEKDYEGKLVLSLOAHKMAWFLFEPVDPNDAPDY 2699
Db 2618 ELIDEEYVCOQOSTEDAMTVLTPLTEKDYEGKLVLSLOAHKMAWFLFEPVDPNDAPDY 2677
QY 2700 YGVIKERMDLATMERVQORRYEKLTEFVADMTKIFDNCRCRYNPSDPFYQCAEVLESFF 2759
Db 2678 YGVIKERMDLATMERVQORRYEKLTEFVADMTKIFDNCRCRYNPSDPFYQCAEVLESFF 2737
QY 2760 YOKLGFKASHSHNNKLOSTAS 2781
Db 2738 YOKLGFKASHSHNNKLOSTAS 2759

RESULT 4
PCT-US01-14827-16042
; Sequence 16042, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; PRIOR FILING DATE: 2001-05-16
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 16042
; LENGTH: 845
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (301)..(316)
; OTHER INFORMATION: PHD-finger domain identified by eMATRIX, accession number
; NAME/KEY: DOMAIN
; LOCATION: (288)..(333)
; OTHER INFORMATION: PHD-finger domain identified by Pfam, accession name PHD, E-
; OTHER INFORMATION: value=4.9e-15, Pfam score of 63.4
PCT-US01-14827-16042

Query Match 23.5%; Score 3363; DB 1; Length 845;
Best Local Similarity 82.6%; Pred. No. 6, 6e-189;
Matches 650; Conservative 4; Mismatches 3; Indels 130; Gaps 2;

QY 1 MVSSEEEEDGDAEFTQDSEDEDEDEDDDDSDYPREMEDDDDDASCTESSFRSHST 60
Db 36 MVSSEEEEDGDAEFTQDSEDEDEDEDEDDDDSDYPREMEDDDDDASCTESSFRSHST 95
QY 61 YSSTPGRRKRVRHRRSPRIIEEKDIPLEPRKSESDIMVNEHIMNVIAYEVLNRNGTV 120
Db 96 YSSTPGRRKRVRHRRSPRIIEEKDIPLEPRKSESDIMVNEHIMNVIAYEVLNRNGTV 155
QY 121 LRLSPFEEDFCALVSOEQCTLMAEMHVVLLKAVLREEDPTMTTGGPADLKDSVNSTLY 180
Db 156 LRLSPFEEDFCALVSOEQCTLMAEMHVVLLKAVLREEDPTMTTGGPADLKDSVNSTLY 215
QY 181 FIDGMTPEVLRVYCESDKERYHNHVLRYQEAEDYRYGVEVNRKIKVLOFLYDQFLTTNARE 240
Db 216 FIDGMTPEVLRVYCESDKERYHNHVLRYQEAEDYRYGVEVNRKIKVLOFLYDQFLTTNARE 275
QY 241 ELMSEGYIQYDDHCRCYCHKIGDLICSETGSAVYHLECVKRPLEVEVDEQCEVCVAHKY 300
Db 276 ELMSEGYIQYDDHCRCYCHKIGDLICSETGSAVYHLECVKRPLEVEVDEQCEVCVAHKY 335
QY 301 PGVTDCAELQOKKRPYIRHNPITYDSSRRKYWFLNRLIIEEDTENENNEKKIYYSGTKYO 360
Db 336 PGVTDCAELQOKKRPYIRHNPITYDSSRRKYWFLNRLIIEEDTENENNEKKIYYSGTKYO 395
QY 361 LAELIDLDLDKYWEALCKILIEEMAREIHRHMDITEDLTNKGASNKSFLLAANEELLES 420
Db 395 LAELIDLDLDKYWEALCKILIEEMAREIHRHMDITEDLTNKGASNKSFLLAANEELLES 460
```



```
|||||
Db 396 LAELIDCDLCKDWEALCKLLEEMKEELHRMDITEDLTNRKAGSKSFLAANEELLES 455
Qy 421 IRAKGGIDNVKSPETERKDKNETENDSKDAEKNREEFEDQJLEKSDDKTPDDPEQCK 480
Db 456 IRAKGGIDNVKSPETERKDKNETENDSKDAEKNREEFEDQJLEKSDDKTPDDPEQCK 515
Qy 481 SE----- 482
Db 516 SEPTPEVGDKGNSVANLGDNTNTATSEETSPSEGRPVGCLSETPDSSNMAEKVASEL 575
Qy 483 ----- 482
Db 576 PODVEEPNKTCESNTSATTTSTIOPNLNENSSSELNSSQESAKAADDPNGEHSHT 635
Qy 483 -----VGFSEKSEKNGELSESPGAGKAGSSTRITIRLNNPSKLSQOYAAA 534
Db 636 PYSIOEIVGDTSEKSTGELSESPGAGKAGSSTRITIRLNNPSKLSQOYAAA 695
Qy 535 HEANKLFEGKEVLVNSOGELISRLSTKEVIMKGNINNYFLGQEGKYRVYHNOYSTNS 594
Db 636 HEANKLFEGKEVLVNSOGELISRLSTKEVIMKGNINNYFLGQEGKYRVYHNOYSTNS 755
Qy 595 FALNKHQREDHDKRRHLAKKCLTPPAGEFKNGSVHGSKVLITSLRITTOLENNIPS 654
Db 756 FALNKHQREDHDKRRHLAKKCLTPPAGEFKNGSVHGSKVLITSLRITTOLENNIPS 811
Qy 655 SELHPNM 661
Db 812 TSLHPSF 818
```

RESULT 5

US-09-629-469A-14273

Sequence 14273, Application US/09629469A

GENERAL INFORMATION:

```
APPLICANT: OTA, TOSHIO
APPLICANT: ISOGAI, TAKAO
APPLICANT: NISHIKAWA, TETSUO
APPLICANT: HAYASHI, KOJI
APPLICANT: SAITO, KAORU
APPLICANT: YAMAMOTO, JUNICHI
APPLICANT: ISHII, SHIZUKO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: WAKAMATSU, AI
APPLICANT: NAGAI, KEIICHI
APPLICANT: OTSUKI, TETSUJI
TITLE OF INVENTION: PRIMERS FOR SYNTHESIZING FULL-LENGTH cDNA AND THEIR USE
FILE REFERENCE: 084335/0123
CURRENT APPLICATION NUMBER: US/09/629, 469A
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: JP 1999-248036
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: JP 1999-300253
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: JP 2000-118776
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: JP 2000-183767
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: JP 2000-241899
PRIOR FILING DATE: 2000-06-09
PRIOR APPLICATION NUMBER: 60/159, 590
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/183, 322
NUMBER OF SEQ ID NOS: 19025
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14273
LENGTH: 557
TYPE: PRT
ORGANISM: Homo sapiens
US-09-629-469A-14273
```

Query Match 18.8%; Score 2695.5; DB 20; Length 557;
Best Local Similarity 79.0%; Pred. No. 8,4e-150;
Matches 553; Conservative 3; Mismatches 1; Indels 143; Gaps 1;

```
Qy 1811 MWAPISGSVYTGCTKMWLTITKVGSPATVTFQOKNKNHOFATVVKOGQSSGVVQOKYL 1870
Db 1 MWAPISGSVYTGCTKMWLTITKVGSPATVTFQOKNKNHOFATVVKOGQSSGVVQOKYL 60
Qy 1871 GLIPSTGTSGOQTFISFOPRTATVIRPNTSSGGGTTSSQVITGPQIRPGMTVIRPLQ 1930
Db 61 GLIPSTGTSGOQTFISFOPRTATVIRPNTSSGGGTTSSQVITGPQIRPGMTVIRPLQ 120
Qy 1931 OSTLGKAIIRTPVWVQPGAPQGVMTQILIRGQPVSTAASAPNTVSSTPGOKSLTSATSTSN 1990
Db 121 OSTLGKAIIRTPVWVQPGAPQGVMTQILIRGQPVSTAASAPNTVSSTPGOKSLTSATSN 180
Qy 1991 IOSSASQPPRPOGGVKLTMAQLTQGHGNGOGITVYIOGQGTGQLDILPGVYVL 2050
Db 181 IOSSASQPPRPOGGVKLTMAQLTQGHGNGOGITVYIOGQGTGQLDILPGVYVL 240
Qy 2051 PPGGQOLMOAMPNGTVORFLFTPLATTATTTTSTTAAGTGEOROSKLSPOKOV 2110
Db 241 PPGGQOLMOAMPNGTVORFLFTPLATTATTTTSTTAAGTGEOROSKLSPOKOV 300
Qy 2111 HODKTLPPAQSSTSVGPAKAQPTAQPASARPOQOTQPSAPQEVQTOPEVQOTTVSSHV 2170
Db 301 HODKTLPPAQSSTSVGPAKAQPTAQPASARPOQOTQPSAPQEVQTOPEVQOTTVSSHV 360
Qy 2171 PSEAOPTHAQSSKPVAAQSQPSQSNVQGSQPVYVQSPQTRIRPSTPSQSLSGQSSQVOT 2230
Db 361 PSEAOPTHAQSSKPVAAQSQPSQSNVQGSQPVYVQSPQTRIRPSTPSQSLSGQSSQVOT 420
Qy 2231 TTSOPRPIQPHNTSLQIPSGQPOSQPOVQSSQTQLSSGGTLMQVSSPSRQQLIQOPQ 2290
Db 421 TTSOPRPIQPHNTSLQIPSGQPOSQPOVQSSQTQLSSGGTLMQVSSPSRQQLIQOPQ 446
Qy 2291 PQYIAPVQLQGVVLSQIQSGVVAQIQAGSQVQQLQLQIQSSAVQTHQIQNV 2350
Db 447 ----- 446
Qy 2351 VTVQAASVQEQLRVQQLRDQOQKKQOQIEIKREHTLOASNOSELQKOVYKKNAYIE 2410
Db 447 -----QVYKKNAYIE 457
Qy 2411 HLKOKKSMTPAERENQRTVCNOYMYTLDKIDKEKOAAKKRRRESVEKRSKONAT 2470
Db 458 HLKOKKSMTPAERENQRTVCNOYMYTLDKIDKEKOAAKKRRRESVEKRSKONAT 517
Qy 2471 KLSALLFKHEQLRAEILKKRALLDKLDIQEVOEELKRD 2510
Db 518 KLSALLFKHEQLRAEILKKRALLDKLDIQEVOEELKRD 557
```

RESULT 6

US-09-614-150-32997

Sequence 32997, Application US/09614150

GENERAL INFORMATION:

```
APPLICANT: Venter, J. Craig
APPLICANT: et al.
TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
TITLE OF INVENTION: DROSOPHILA GENES.
FILE REFERENCE: C1000728
CURRENT APPLICATION NUMBER: US/09/614, 150
CURRENT FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/157, 832
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: 60/160, 191
PRIOR FILING DATE: 1999-10-19
PRIOR APPLICATION NUMBER: 60/161, 932
PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: 60/164, 769
PRIOR FILING DATE: 1999-11-12
```



```

; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 32997
; LENGTH: 976
; TYPE: PRF
; ORGANISM: DROSOPHILA
US-09-614-150-32997

```

```

Query Match      10.9%; Score 1557.5; DB 20; Length 976;
Best Local Similarity 36.4%; Pred. No. 3.1e-82;
Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;

```

```

QY 4 EEEEDGGALETTODSEDEDEEDDDSDVPEEMEDDDDDASYCTESSFSHTYSS 63
DB 99 ESEYHYGSDGDESEEDKSDNEDMDLLTPSDDESLEAVANESESEFSVC---SFNONGV--- 152
QY 64 TPGRRKPRVHRPSPILIE-KDIPLEFPKSEDLVWPNHINVAIYEVLRNFGTVLR 122
DB 153 ---GRRPRPSPVWMLQEGROYAALDLPDSSDDLFIANTHVLRALSIYELRRFRHMVR 209
QY 123 LSPFRREDFCAALVSQOCTLAEMHNVLLKAVLRBEDTSNTTFGRADLKDSVNSTLYFT 182
DB 210 LSPFRREDFCAALACEQSALTEVHIMLKALIREDDAGTFFGRPLDQDVTYNISLYLI 269
QY 183 DGMTPEVLRVYCESKEY----NHVLPYOEAADYPRGPEENKIKYLOFLVDOFLTTNIA 238
DB 270 DSTTPEVLRVSYESDKTEPRANVFHLSHTE---YPTGIDNLEVLQFISDPLFISNST 326
QY 239 REELMSEGVLYQDDHCRVCHKLDDLCCETCSAVHLECYKPRLEEVPEDEMOECVAVH 298
DB 327 RDVMLQEGPIHYDDHCRVCHRLDGLCCETCPAYVHLECVDPMDNVPREDMOCGLCRSH 386
QY 299 KVGCVTDCAVALEIOKNRPYIRHPIRGYDRSRKRYWFLNRRLIEEDTENEKEKIWYYSRK 358
DB 387 KVGCVTDCAVALEIOKNRPYIRHPIRGYDRSRKRYWFLNRRLIEEDTENEKEKIWYYSRK 442
QY 359 VOLAEILDCDKYWEAELOKLEEMREETHRMIDTEDITNKRASGNSKFLAANDEIL 418
DB 443 SKKLILSRDALELETRLHSQTERDELEROMKLETELTINHKHTKRSYI----- 494
QY 419 ESIRAKGIDINVKSPEETEKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDDEQ 478
DB 495 -----ETEOEAKN-----ELLEKEVLDEDEKD----- 516
QY 479 GKSEVDFKSEKNGELSESPGAKGASGSTRITRLRNDSLSQKSOQVAAAHAHEAN 538
DB 517 -----GDASESOSIE-----GTKKOECKKWTYR-----OKSQOLT----- 547
QY 539 KLFKEGSEVLYVNSOGEISRLSTKKEVIMKGINNINNYFKLGOEGKYRYVHNOYSTNSFALN 598
DB 548 -----NCTLHFKLGMEOGFKNYVNOYSTNPLALN 576
QY 599 KHOHREDHOKRRLAIFCLTPAGEFRKWSVHSGVLYLTSTLRLLTYLOLENNIPSSFLH 658
DB 577 KPGORNEEROKRRHLSHFSTLTASDFKWTIGITGTDNMITTLRQTLINESNIASFLN 636
QY 659 PMAASHRAVNIKAIVOMCSREFALALALECAVCPVYMLPIWRFLGHRLLRMSIER 718
DB 637 INNVVAKKILNAAVMAARRSEFAVLLLOASLKSVPANVHNOJGHTTLQRLITISAER 656
QY 719 EEKEKVKKKEKKO---EEETMOQATVWKKYTPVVKHQVWKQKGEERYVIGYGWSMISTK 775
DB 697 EEKRLKLEKREKRDEEENRRLAFNVIKYTLGLKHQVWKQKGEERYVHNOJGHTTLQRLITISAER 756
QY 776 -----HYVIRVPEKLPQNTNVNTRKSLDGTKNMDENNDESDDRKCSRPKIKIIEPDSEK 830

```

```

DB 757 RRCGVARRAQAQPLTHNRVYHYTM---GENDVNEIT-----LVDPRTQRFMOOCES 805
QY 831 DEYKSGDAKAGAO--NEMDISKITEKKODVYVELLDSDSDKCRKEPHEVDDMTESH 888
DB 806 SNVDGVCYCHLPQYKRVKVKIEDVTE-----KIKGH 836
QY 889 VNCQESSQVNVVAVSGEHL--RTSYKKTKTSKLDGLERIKOTFLKEORLEKIKLE 946
DB 837 -----IDYSKALNAPGRTIYSKVAARKSRDLDDLRLKLALEVEBOM---ASKIT 882
QY 947 GGIKGIGKTSTNSKNLSSES 966
DB 883 SDMKPILVSSQNNNTANSKOT 902

```

```

RESULT 7
US-60-191-637-32572
; Sequence 32572, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; FILE REFERENCE: CL0000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 32572
; LENGTH: 976
; TYPE: PRF
; ORGANISM: DROSOPHILA
US-60-191-637-32572

```

```

Query Match      10.9%; Score 1557.5; DB 27; Length 976;
Best Local Similarity 36.4%; Pred. No. 3.1e-82;
Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;

```

```

QY 4 EEEEDGGALETTODSEDEDEEDDDSDVPEEMEDDDDDASYCTESSFSHTYSS 63
DB 99 ESEYHYGSDGDESEEDKSDNEDMDLLTPSDDESLEAVANESESEFSVC---SFNONGV--- 152
QY 64 TPGRRKPRVHRPSPILIE-KDIPLEFPKSEDLVWPNHINVAIYEVLRNFGTVLR 122
DB 153 ---GRRPRPSPVWMLQEGROYAALDLPDSSDDLFIANTHVLRALSIYELRRFRHMVR 209
QY 123 LSPFRREDFCAALVSQOCTLAEMHNVLLKAVLRBEDTSNTTFGRADLKDSVNSTLYFT 182
DB 210 LSPFRREDFCAALACEQSALTEVHIMLKALIREDDAGTFFGRPLDQDVTYNISLYLI 269
QY 183 DGMTPEVLRVYCESKEY----NHVLPYOEAADYPRGPEENKIKYLOFLVDOFLTTNIA 238
DB 270 DSTTPEVLRVSYESDKTEPRANVFHLSHTE---YPTGIDNLEVLQFISDPLFISNST 326
QY 239 REELMSEGVLYQDDHCRVCHKLDDLCCETCSAVHLECYKPRLEEVPEDEMOECVAVH 298
DB 327 RDVMLQEGPIHYDDHCRVCHRLDGLCCETCPAYVHLECVDPMDNVPREDMOCGLCRSH 386
QY 299 KVGCVTDCAVALEIOKNRPYIRHPIRGYDRSRKRYWFLNRRLIEEDTENEKEKIWYYSRK 358
DB 387 KVGCVTDCAVALEIOKNRPYIRHPIRGYDRSRKRYWFLNRRLIEEDTENEKEKIWYYSRK 442
QY 359 VOLAEILDCDKYWEAELOKLEEMREETHRMIDTEDITNKRASGNSKFLAANDEIL 418
DB 443 SKKLILSRDALELETRLHSQTERDELEROMKLETELTINHKHTKRSYI----- 494
QY 419 ESIRAKGIDINVKSPEETEKDNETENDSKDAEKNREEFEDOSLEKSDDKTPDDDEQ 478
DB 495 -----ETEOEAKN-----ELLEKEVLDEDEKD----- 516
QY 479 GKSEVDFKSEKNGELSESPGAKGASGSTRITRLRNDSLSQKSOQVAAAHAHEAN 538

```



```
Db 517 -----GDAKSEOSIE-----GTRKQEBCKMVT-----OKSNQLT----- 547
Qy 539 KLFKEGKVLVYNSGGEISRLSTKKEVIMKGNINNYEKLQEGKRYVYHNOYSTNSPALN 598
Db 548 -----NOTLHFKLQMEGCFKNYNOYSTNFIALN 576
Qy 599 KQOHRDHDKRRHLAHKFLTPAGEFRKNGSVHGSVLTISTRLTLITOLENNIPSSFLH 658
Db 577 KQORNEERDKRRHLSHKFLTTSADPKWIGITMGTDNMITTLRQTLINFEINIASFLN 636
Qy 659 PNMASHRAWIKAVQMSCPREFALALALIECAVPRVYMLPIREELGTRHLRHMSIER 718
Db 637 IMVYVKKKIMNAAVMNAARPSSEFAYVLLLPQASLKSVEFANVWHEQGLTLORITSAER 696
Qy 719 EERKVKKKKKKKO---EEBETMQATWVKTTPVVKHQVKKQKEEYRVYGGSMVSKT 775
Db 697 EERKLEKREKREDEBERNRLAFNYIKYTLGLKHQVKKQKEEYRVHGSQWMLWSSS 756
Qy 776 -----HYFVFKPLPGNTVNVYTRKSLGTRKNMNMENMDESCKRCSRPKKIKIEPDESK 830
Db 757 RRCGVARRAQPLTHNRVYVHYTM---GEENDVNEIIT---LVDPRTQRFMOQCES 805
Qy 831 DEYKSDAAKADQ---NEMDISKITEKKDQVYKELSDSDCKPEPEVDDMKTESH 888
Db 806 SNVDGQVCHYLPDQYKNVYIEDVTE-----KIKGH 836
Qy 889 VNCQESSQVYVNVSEGFHL--RTSYKKTKSSKLDGLERIKOTPLEKORLEKIKLE 946
Db 837 -----IDVSKALNAPGRTYYSKVARKSRLDLDLRKLAEVEBOM---ASKIP 882
Qy 947 GGIGIGIKTSTNSSKNLS 966
Db 883 SDMKPLVSSQNNNTANSKQT 902

RESULT 8
US-60-191-681-25768
; Sequence 25768, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: LI, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25768
; LENGTH: 976
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-681-25768

Query Match 10.9%; Score 1557.5; DB 27; Length 976;
Best Local Similarity 36.4%; Pred. No. 3,1e-82;
Matches 357; Conservative 138; Mismatches 292; Indels 193; Gaps 22;
```

```
Qy 239 REELMSEGVLYOYDHCRCVCHKLDLLCCETCSAVYHLECYKPPLEVEPDEMOCEVCVAH 298
Db 327 RDVMLQEGPIHYHDHCRCVCHRLGDLCCETCPAVYHLECYDPPMNDVPTEDMCGCLCRSH 386
Qy 299 KQVGVYDCAEIOKKNPYTRHEPIGYDRSRKRYFLNRLIIEEDLENEEKKIWTYSK 358
Db 387 KQGVYDVCVLPQKQGVLLRHDSLGVRHGRKRYWFLARLFI--EDENFNP---CWYSTT 442
Qy 359 VOLAEILDCDKDYWEALCKIIEEMREIHRMDTDEDTLNKARGSNKSFLAANBEIL 418
Db 443 SKLLLSRLDAELETRLHSQITERDETEROMKLETLTNHKKTKSVI----- 494
Qy 419 ESIRAKGIDIDNVKSPETEKKNETENDSKDAEKNREEFDSLKSDSDKTPDDPEQ 478
Db 495 -----EIEQEAKN-----ELLEKVLDEDEK----- 516
Qy 479 GKSEVDGFKSEKNGELSESPAGKASGSTRITTLRNPDSKLSQKQVAAAHAN 538
Db 517 -----GDAKSEOSIE-----GTRKQEBCKMVT-----OKSNQLT----- 547
Qy 539 KLFKEGKVLVYNSGGEISRLSTKKEVIMKGNINNYEKLQEGKRYVYHNOYSTNSPALN 598
Db 548 -----NOTLHFKLQMEGCFKNYNOYSTNFIALN 576
Qy 599 KQOHRDHDKRRHLAHKFLTPAGEFRKNGSVHGSVLTISTRLTLITOLENNIPSSFLH 658
Db 577 KQORNEERDKRRHLSHKFLTTSADPKWIGITMGTDNMITTLRQTLINFEINIASFLN 636
Qy 659 PNMASHRAWIKAVQMSCPREFALALALIECAVPRVYMLPIREELGTRHLRHMSIER 718
Db 637 IMVYVKKKIMNAAVMNAARPSSEFAYVLLLPQASLKSVEFANVWHEQGLTLORITSAER 696
Qy 719 EERKVKKKKKKKO---EEBETMQATWVKTTPVVKHQVKKQKEEYRVYGGSMVSKT 775
Db 697 EERKLEKREKREDEBERNRLAFNYIKYTLGLKHQVKKQKEEYRVHGSQWMLWSSS 756
Qy 776 -----HYFVFKPLPGNTVNVYTRKSLGTRKNMNMENMDESCKRCSRPKKIKIEPDESK 830
Db 757 RRCGVARRAQPLTHNRVYVHYTM---GEENDVNEIIT---LVDPRTQRFMOQCES 805
Qy 831 DEYKSDAAKADQ---NEMDISKITEKKDQVYKELSDSDCKPEPEVDDMKTESH 888
Db 806 SNVDGQVCHYLPDQYKNVYIEDVTE-----KIKGH 836
Qy 889 VNCQESSQVYVNVSEGFHL--RTSYKKTKSSKLDGLERIKOTPLEKORLEKIKLE 946
Db 837 -----IDVSKALNAPGRTYYSKVARKSRLDLDLRKLAEVEBOM---ASKIP 882
Qy 947 GGIGIGIKTSTNSSKNLS 966
Db 883 SDMKPLVSSQNNNTANSKQT 902

RESULT 9
US-60-173-464-22080
; Sequence 22080, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: LI, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22080
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
```


LOCATION: (1)...(786)
OTHER INFORMATION: Xaa = Any Amino Acid
US-60-173-464-22080

Query Match 10.4%; Score 1493; DB 27; Length 786;
Best Local Similarity 40.5%; Pred. No. 1.4e-78;
Matches 316; Conservative 110; Mismatches 224; Indels 130; Gaps 14;

```
QY 4 EEEEDGDAEETODSDDEDEMEDEDDDDSDYPEMEDEDDDDASCTESFSRSHYSS 63
  | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 ESEHYGSDGDESDKSDNEDMDLLTPSDDELEVAANESESEFVC---SFQNGV--- 131
QY 64 TPGRRPRVHRPSPILEE-KDIPPLEPKSSDLMVPHNHNVAIYEVLRNFGTVLR 122
  | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 ---GPRPRPSPEPVMLQEGROQYALDLPDSSDGLFANTHVALISTYEVLRFRIMVR 208
QY 123 LSPFREDFCAALYSOECTLMAEMHVVLLKAVLRREDTSNTTFGPADLKDSVNSTLYFT 182
  | | | | | | | | | : : : : : : : : : : : : : : : : : :
Db 209 LSPFREDLCAALACEGOSALTEVHIMLKLKALRGEDAGTFHFGPLDQDTYNSILYLI 268
QY 183 DGMTPEVLRVYCESDKEY---HHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTNIA 238
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 269 DSTTPEVLRVYSESDKTFPRNVPFHLSITE---YPTGIDNLEVLQFLSDQFLSNSI 325
QY 239 REELMSEGVLYDDHCRVCHKLGLDLCETCSAVYHLECVKPLEEVEDEMOECVAVH 298
  | : : | | | | | | | | | : | : | : | : | : | : | : | : | : | :
Db 326 RDVWLQGPRIHYDDHCRVCHRLGLDLCETCPAVYHLECVDPMDNDVPTEDMOGCLCRSH 385
QY 299 KAVGVYDCAVEIQRNRYTHPEPIGYDRSRKRYWFLNRLLIEEDTPENEKKIWTYSTK 358
  | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 386 KAVGVYDCAVLPQERQVGLIHSDISGVDRHGRKYWFARLFTI-EDQENFT---CWYSTT 441
QY 359 VOAEELIDCKDYWEALCKILEEMREELIHRHMDITFEDTLNARGSKNSFLAANEEL 418
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 442 SKTLULSRDABELETRLSQITERDELIERQMKLFETITLNEKHKTRSVI----- 433
QY 419 ESTRAKKGIDINVKSPETEKKDNETENDSKDAKNREEDQSLKSDSKTPDDDEQ 478
  | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 494 -----ELIQEAKN-----ELLEKEVLDEDEK----- 515
QY 479 GKSEVGDGKSEKNGELSESFGAGKSGSTRITLRLRNPDSKLSQSOOVAANHEAN 538
  | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 516 -----GNAKSESOSIE-----GTKKQEBCKWATR-----OKSMOLT----- 546
QY 539 KLFKEGKEVLVNSOGELISRLSTKKEYIMKGINNYPFKLGOEGKYRVYHNOYSTNFPALN 598
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 547 -----NGTLHFKLMEGQFKYVNOYSTNFPALN 575
QY 599 KHQREDHDKRRHLAKFCCLTPAGEEFKNGSVHGSKVLITSTLRITITQLENNIPSSFLH 638
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 576 KQORNEERDKRRHLSHFSSTLTASDFKWIITGCTDNMTTLRQTLINFESNIAASFLN 635
QY 659 PNMASHANMIKAVQMSKREFALALILCAVKKPVVMLPIREFLGHRLRHMSTIER 718
  | : : : : | : : : : | : : : : | : : : : | : : : : | : : : :
Db 636 INNVVAKKINNAAMARRSEFAVLLLFQASLSKSYFANVMEHQGHITTLQRTITSYER 695
QY 719 EEKEKVKKKKKQ---EEETMQOATVKKYTPYKHQVOMKQGESEVYTGGSWISKT 775
  | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 696 EEKKLEKREKREDEENRRLAFNIKTYTLGLKHQVWKQGESEYVHOMQWMLSSS 755
```

RESULT 10
US-10-029-386-32957
Sequence 32957, Application US/10029386
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029, 386
CURRENT FILING DATE: 2001-12-20

```
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 32957  
LENGTH: 273  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC023123.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.8  
OTHER INFORMATION: SWISSPROT HIT: Q12830, EVALU0.00e+00  
US-10-029-386-32957  
Query Match 10.3%; Score 1478; DB 24; Length 273;  
Best Local Similarity 100.0%; Pred. No. 2e-78; Indels 0; Gaps 0;  
Matches 273; Conservative 0; Mismatches 0;
```

```
QY 67 RRRPRVHRPSPILEEKDIPPLEPKSSDLMVPHNHNVAIYEVLRNFGTVLRSPF 126
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 RRRPRVHRPSPILEEKDIPPLEPKSSDLMVPHNHNVAIYEVLRNFGTVLRSPF 60
QY 127 REEDFCAALYSOECTLMAEMHVVLLKAVLRREDTSNTTFGPADLKDSVNSTLYFTDGMT 186
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 REEDFCAALYSOECTLMAEMHVVLLKAVLRREDTSNTTFGPADLKDSVNSTLYFTDGMT 120
QY 187 WPEVLRVYCESDKEYHHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTNIAEELMSEG 246
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 WPEVLRVYCESDKEYHHVLPYQEAEDYPGVPEKIKVLOFLVDOFLTNIAEELMSEG 180
QY 247 VIQYDDHCRVCHKLGLDLCETCSAVYHLECVKPLEEVEDEMOECVAVHKGVTDC 306
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 VIQYDDHCRVCHKLGLDLCETCSAVYHLECVKPLEEVEDEMOECVAVHKGVTDC 240
QY 307 VAEIQKKNRYTHPEPIGYDRSRKRYWFLNRLLI 339
  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 VAEIQKKNRYTHPEPIGYDRSRKRYWFLNRLLI 273
```

RESULT 11
PCT-US00-05989-816
Sequence 816, Application PC/TUS0005989
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Human Pancreas and Pancreatic Cancer Associated Gene Sequences
FILE REFERENCE: PA105PCT
CURRENT APPLICATION NUMBER: PCT/US00/05989
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 928
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 816
LENGTH: 328
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (170)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (172)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (174)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE


```

Db 449 RNMATTEASSAAKALNVQPTDNSFODVIDYNGD---TKVMTVYAGQTFTR----- 497
QY 656 FLHNMAASHANWKAQOMSKPREFALALILECAVNPVMLPIMGEFLGHTRLHMTS 715
Db 498 -----NLTWTKN---SGCTFLSLMSTASTGCAKNLOVOFGTFEYESAVAKVRY 545
QY 716 IEREK-KEKYKKKKKEEETMOATVWVYTFEPVKHOV--WKQGEHEVYT-----GY 766
Db 546 VDANTGDIIPKTIAGEVATV-----IDKQLNNKNSGYSYVSTALQNSNY 595
QY 767 GGSMT-----SKTHVREVPKLPGNVNTWRKSLSECTKANMDEMSDKRCSRS 818
Db 596 SETSGPTLKLITNSQVIVIKF-----KDVOGPOISVD-----SQIREGKT 637
QY 819 PKRKI-EPSEKD---EYKSDAAKAGO-----NEMDISKITEKKDDVVEL 864
Db 638 INPTITTTDMSKDVLTYYTGLPSGLSFDOTNTITGTSEVGTIVT-----VNTT 690
QY 865 DSDSDKCEPEMEVDMDKTESHVNCQESSOV-----DVVNS 903
Db 691 DARGNVTSKQFTITIOPTISPVVWTFPSQASEVFTPINPTITATDMSGVVTFVTGLP 750
QY 904 EGFILRTSYKKTKSSKLDLERRIKOFTLEKQLEKIKLEGIGIKGTSTNSKNL 963
Db 751 QGLFADASTNSIVGTPQIGNTITITESTDASGNKTKTKINYE-----VTRNSASDS 802
QY 964 SESPVITRAKEGCO---SDSMROEOSPANANDQPEDLLQOGCSOSVLRMSDPGHTT 1018
Db 803 TSTIVASVSTNSNSTSLSDSVKASOSLSTKSLSESL--SASTNSSTISQASESASTS 860
QY 1019 NKLYPKRDVLADVINSRPTKCPKQNSIENDIEKVSDLASRGOEPTKSTKGNDFIDD 1078
Db 861 KOL-----SESASTSTSDASASARKESTKSTKS-----LSE 893
QY 1079 SKLASADIGLICKNNKPLIQESDPTIVASSKALHSSVPKSNDDARPILSRAMFEG 1138
Db 894 STSTSVSDASAVSTSES---ASTSTSVSGSTSTSTSTSTSTSTSDASIKASASAS-TS 948
QY 1139 KLGCDSNSNSTLENSSDTVASIODS-SEEDMIVONSNEISIEFOFTRQDOVEVELEPKCEL 1197
Db 949 KILSESVSTSTSDASASTSTSVSDNSASTSLKSTSTSVSDSTSTSD----- 997
QY 1198 VSGSTGNCEDRLPVKGTENGAKKPSOOKLIEEPVVKCSDQIKLAKTTDKNNENRESE 1257
Db 998 --SASTSTSE-----SESDSASTSLSESTSTSVSDSTSTSTSDASASASESE 1043
QY 1258 KKGORTSTFOINKDKNPKIYIKGECKEISESRVSGNEVPKNVNNIKIIPENDIKSLT 1317
Db 1044 SNKSTSTST-----LSF-STSTSLSGSTSTAST-----SDSASTS 1074
QY 1318 VKESAIRPFINGDIVIMEDFERNSSETKSHLSSSDAEGNYRDSLETLPTKESDSTQTT 1377
Db 1075 TSESE-----SDSTSTSLSESTSTSLSGS-TSASTSDAST--STSESDSTSES 1120
QY 1378 TPASACSESNVNOVEIMEIFTSYKAVTSPITSEESN-----LSNPTIDENGL 1428
Db 1121 T--SLSESLSTSVSDSTSTASTSE---SASTSTSEESNSASTSLSGSSTSTSTSTST 1173
QY 1429 PINKENUNV-GESEKRTVITEVTMTSTVAETSEKTVIKVEKGQKQTVVASTENCAKTVT 1487
Db 1174 STSDASTSTSESDSTSTSTSTSTSTSTSTSTSTSEASASTSTSESDSTSTSTSLSE 1233
QY 1488 TTTTTVTKLSTPSTGSDVDI--ISVKEQSKTVV-----TTTVDLSLTGTLVMTVS 1540
Db 1234 STSTSVSDSTSTASTSDASTSTSTSVSDSEASSTSTSESLSTSVSDS-TSTSTSDASASTS 1292
QY 1541 KEYSRDKVILMKFSRKRKTRTSALTYSYKFTYKSTKSKSLFVLPNDLKLAKGCIRE 1600
Db 1293 ESDSTSESTSLSESTISVSDSTSA--STSDASTSTSES-----ESDSASTSLSG--- 1341
QY 1601 VPPYNNAKPALDIWPPSPRPFTGIIWRYRLQTVKSLAGVSLMLRLMLMSLRMDDMAK 1660
Db 1342 -----STSTSLST-----STSTSLST----- 1348

```

```

QY 1661 VPPGGCSTRTETSETTEITTEIIRKRDVGPYGIREFEXCIKRIICPIGVETPEKTEPTQR 1720
Db 1349 -----DSTSTSDASASTS-----ESDSERASTSL 1375
QY 1721 KGLRSSALRPKRPEPTKQTPV-----IETVABEELMELMIRAFAER 1764
Db 1376 SGSTSTSLSDSTSTSTSTSDASTSTSVSDNSASTSLSGSTSTSTSTSTSTSTSDAST 1435
QY 1765 VEKEMKAVQOAKKRLQOQKPTVIATSTSPSTSTSTSPROKVMVAPISGVTTGKT 1824
Db 1436 SEDD-----SERASTSLSGSTSTISDSSTSTSDASASTSVSE-----SNSTSTIS 1484
QY 1825 MVLTKVGSPTAVTPOONKNEHOTFATVWQOGSNGSVVQOQVGLGIIIPSTSQOQTE 1884
Db 1485 ESLSSTSV-SDSTST-----STSDASTSTSVSDSDASTSSSSSV-----STSDSEST 1532
QY 1885 TSFOPTAIVTI-----RPTSGSGGTTSSQVITGPQIRPQMTVIRTPLOQSTLGKAIT 1939
Db 1533 TSTSDASTSTSVSESNSTSTSTSLSGSTSTSVSDSTSTSTSDAS---ASTSESDSDAST 1589
QY 1940 RTPVMVQPGAPQOQVMTQILINGQPVSTAVSAPNTVSSNPGOK---SLTSASTSTSIQSSAS 1996
Db 1650 STSESDSDASTSLSDSTSTSVSESTSTSTSTSVSASNSTSTSLST--DSRSTSLSDSTST 1707
QY 2057 LMOQAMPNGIVQRLPPLATTTATTTTYYSTTRAGIGEQRQKSLSTQOMVHQDKTL 2116
Db 1708 STSESGSTSTSES--DSDASTSLSESTSTSTSTSTSTSTSTSDASTSTSVSDNSASTSL 1765
QY 2117 PPAOSSSV--GPAKAOPTQAPARPOPTQOPSPAPQPEVOTQPTVTSVSHVSEA 2174
Db 1766 SDSTSTSVSDSTASTSEASASTSTRESESTASTSLSESTSTSVSDSTSTSD---SAS 1822
QY 2175 QPTHAOSKRPQVAAQSPQSNVQOQSPVAVQSPQTRIRPSTPQSLSPGQOQVQTTTQ 2234
Db 1823 TSTSESDNSNESTSLSESTSTSTSTSTSTSTSTSTSTSTSTSTSDASTSTSVSDNSASTSLSGST 1877
QY 2235 PIRIPIHTSIQIRPQOQPOQPOVQOSTOTLSSGOTLNOVSVSPSRPQLOIQPOQV 2294
Db 1878 STSVSDSTSTSTSTSTSTSTSTSESDSDASTSLSGSTSTSTSTSTSTSDASTSTSE-- 1935
QY 2295 AVPOLQOQOVLSQIQSOVVAQIQAOQSGVPOQIKLOLPTQIOQSSAVQTHQIONVTVQ 2354
Db 1936 -----SASTSTSVSESDSESTSVSESS-----TSVSDSTSTSTSEASASTSTSE 1979
QY 2355 AASYOQLOQVQOLRDQOQKKKQOQIEIKREHPLQASNOSEIIOQOVYMKHNAVIEHLQ 2414
Db 1980 SESTSES-----TSVSESSSTSTISDSSSTSTSMSTSEFTSQSPINSSEQIFGLSLE 2032
QY 2415 KKSMTFARENOMIYONVMKYILDKIKKEQOAKKRRREBSVQKRSQKNATKLSA 2474
Db 2033 DTIVT--QSKNTNML--NKTGK---DYDLOQKGYTDBQHNHETQSNQADHNSNLDL 2083
QY 2475 LFEKHEQLRAEILKK 2490
Db 2084 L--HQRNRLODKVVKQ 2096

```

```

RESULT 2
US-10-092-411A-5080
; Sequence 5080, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001

```


Db 2182 ----GQGRSTIDQNMATTEKEKQALERLNOETNGVNDRIQALANONVTDEKNILLETI 2237
Qy 2079 ATT-----ASTTTVTSTTAAGTGEOBOSKLSPOQVQODKTLPPAOS 2121
Db 2238 RNVEPIVIVKRAKEIIRKRAAEQTLINONQDATTLEKQIALGLKEVRNEALNOVSOA 2297
Qy 2122 SSVPAPK-AOPOTQPSARPQOTOPQSPAPQEVQ--TOPEVOT-----2162
Db 2298 HSNNDVIAEENGLAKISEVHEPETIIRKNAKQETEDQAOQIDITINANKNSTNEKSAI 2357
Qy 2163 -----QTVSSHVPSQADPTHAOS-----2182
Db 2358 DRVNVAKIDALINNTATTTQLVNDAKNSGNTSISQILPSTAVKTNALALASEAKKNNA 2417
Qy 2183 ----KPOVAOSQPSN-----VQOSPVRYQSPS 2208
Db 2418 IIDOTPNATAEKEEANKKYDRLOEADANILKAHTTDEVNNIKNAVONINAVQEVIK 2477
Qy 2209 QTRIRPSTPQOLSQSOVQTTTSCPIQPIQPHSLQIPSGQOPQSPQVQOSTOTLSG 2268
Db 2478 KQNVK-NQNLQFIDNQKLIET-----PDATLEKKAENRLQNLVLTSTDEIANY 2528
Qy 2269 QTLNQSVS-SESRQOLQIQOPQVAVAVQVLOQOVQVLSQVQVQAQ--IOAQSGV 2324
Db 2529 DHNEVDQALDKARPKIE-----ELVPQVSKRDVLNAIQEAFNSQOQELQENQEPAT 2580
Qy 2325 PQOLKLOLQIQIOO--SSAVQTHQIQNVVTVQASVQEOQLQRYQOLRDOQKQKQOQIE 2381
Db 2581 NEE-KTEALNKINQNLNQAKNIDQOSKNKVDQSAKTR-STQDEIQIQPHQTK-----A 2633
Qy 2382 IKREHTQASNOSETIQOVVMKH-NAVIEHLKOKSKMTPAEREENQMIYCNQVMKYIL 2440
Db 2634 TGRHRLNEKANQ-----QOSTIATHPNSTIE-----EROEASAKL--QEVLEKAI 2676
Qy 2441 DKIKKEQAAKKRRRESVQOKRSKQVATKSLALFKHKEQLAEILKKRALLDKDIOI 2500
Db 2677 AKIDGQNDVDEKTVVNGVIAEIEILPATTV-----KKAADVAQAEK--QKNIQI 2727
Qy 2501 EVOEELKBDLKKREKIDLMQAOATAVAAPCPYTPVLPAAPAPPPPPGVQHTGL 2560
Db 2728 NSNDQATTEEKLVASDNLNHVVEETTQALIEDAPDINOQVNE-----KNGCIG 2774
Qy 2561 STPLPVAQOKRRKEEDSSSKSKKMISTSKETKKDK 2602
Db 2775 TIRDIQPLVKKPTAKSKIESAVEKKEIKTEINQTONATHEVR 2816

RESULT 3
US-09-724-676-65945
; Sequence 65945, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 65945
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-65945

Query Match 2.8%; Score 401; DB 5; Length 3256;
Best Local Similarity 18.7%; Pred. No. 0.005;
Matches 527; Conservative 364; Mismatches 1016; Indels 912; Gaps 131;

Qy 339 IIEEDTENENEKKIWTYSTKVQALAILDCLDK---DYWEALCKIILEMREELIHRNDI 394
Db 432 VLPETEIHNPEFLTLMLOYERKIQKDSLSKPEKLTGTAGQCMGSLPGLSS-----VDI 486
Qy 395 TE-DLTNKARG-----SNKSFLLAANEIILE-----SIRAKGDIDN-----VKSPEE 436

Db 487 NNFQDSINESEGIPLKRKRVSFGHLRPELFDENLPNTPLKRGAPTRKSLVMTTPPV 546
Qy 437 TEKDKMETENDSDAKENR---EFEDQSL-----ENDSDQKPR-DDDPPQSGSEVDFKS 488
Db 547 LKRTIKQPPQSGQSGSEIHVEVAQSLVSPAPSPKTFVASDQRRSCRTAPASS 606
Qy 489 EKSNGEISESPGAKGASGSTRILITRLRNPD-----KLSQLKSOQVAAAHEANKLR 542
Db 607 SKQOTVPRK-----RGERVATCLOKRVLSISRSQOHLDMICKSRSGASEAN-----654
Qy 543 EGKEVLVNSQGEISRLSTKK---EYIMKG--NINNYKLGQEGKYRV--YHNQYSTN- 593
Db 655 ----LIYAKSMADYVVKGAQOTQTKVIKHPQKRSNMKRRRPATPKRPVGEVHSQFSTGH 710
Qy 594 ----SPALMKHQRHEDHKRR--HLAHRKCLTPAGEFK-----WNGSVHSGKVLPT 637
Db 711 ANSPCTIIGKAHTEKHVPARYRVNLNFTSQKMDFKEDLSGLIEMKTPYKQEPQLT 770
Qy 638 ISTLRITITOLENNIPSSF-----LHPNASHRANIKAVQCMCK-PREFALALAI 687
Db 771 -STCHIALSSENILGQFGQDTSGEERLLPTSESGVNFESAQNAKOPSD-----822
Qy 688 LECAYKPVWMLPIWREFLGH-----RLHMTSTIER-----719
Db 823 -KCSASP-----PLRQOCIRENGVNAKTPRNTYMTSLETITSTETEPKTYSTVNRSGR 877
Qy 720 --EKEVKR--KKEKQEEETM-----QOATWKKYTFPKQHVQMKQGEYVATGV 766
Db 878 STERNIQKLPVHKSSEETETELVEICLRKGQATLLQO-----RREGEMKEI---925
Qy 767 GQMSWISKTHVYRPVKLPQNTVNVNRKSLGTRKNNDENMDSK-----RKCS--RSP 819
Db 926 -----ERPPEET--YKENIETLEN--DEMKKMKRSRTWGQACAPMSDL 964
Qy 820 KKIIEPDS--KDEVKG-----SDAKGADQNDMDISITEKKQDVYKELDSDSKP 871
Db 965 TDLKSLPDTLMKDTAKGQNLQTDHAKAPKSEKGIITM-----P 1006
Qy 872 CKE-EPMEVDDDMKTESHVQCOESSQVYVNVSE--GFHLRTS-----911
Db 1007 QQSLQF--EPINTPHTHQQLKASLGKGVKEELLAVGKFTRTSGETHTHREPAQDG 1062
Qy 912 ----YKKKTK-----SSKIDGLIERIKOFTLEKQRLKILBGIKIGTSTNSSK 961
Db 1063 KSIKTFKESQIQLIDPARAYTGM--KKWPTPKREAOQLDLDL--AGFKELFTQPGSEE 1117
Qy 962 NLSEPVYITRAKEGCOSDMSRQEOSPMANNDOPE-DLQGCSCQSDSSVLRMSDPSTTNK 1020
Db 1118 SMIDEKT---TKIACKSPPEESVDTPTSTKQWPKRSIRKADVEEELALAKLTPSAGKAM 1174
Qy 1021 LYPKDRVLDDVSI-----RSPETKCPKONSTEN-----DIE 1051
Db 1175 LTPRPAGDEKDIKAFMGTPVQKIDLAGTLPGSKRQLOTPREKAQALIEDLAGKELFQTP 1234
Qy 1052 EKVDLDSRQ-----EPTSKTGTGNFFLDDSKLASADDIGTLI--CKNK 1095
Db 1235 GHTBELVAAGKTKRIKPDSPQSDPVDTPTSTKQK-----PKRSIRKADVEGELALAKRL 1288
Qy 1096 KPLIOESDRTVSSSKSALHSSVPKSTNDRD-----ATPLSRAMPDEGKLGCDSESNSTL 1150
Db 1289 MP-----SAGKAMHTPKPSVGEKDIILFVETPVQKL-----DLT 1323
Qy 1151 ENSSDVTVSIODSSEDMIVONSNESISEQPTREQDVLEPLKCELVGSESTG-NCEDR 1209
Db 1324 ENLGSKRPRQPTKPEEAQALIEDLIGFKELFQTPGHTEEA-----VAAKTKTKMPCSS 1376
Qy 1210 LPVGTGANGKPKSQOKLEERPVYKNSQDQIKLNTDKNNRESEKKQGOPTSPQIN 1269
Db 1377 PPEASDPTSTTRQPKPILEKRDVQKELSAK-----KLQTSGETTTHDKVP 1424
Qy 1270 GKDKPKIYLGCELTKEISRVVSGNV-EPKVNINIKIIPENDIKSLTVKESAIKRPIN 1328


```
Db 1425 GGEKDSINAFRETAOKIDLPASVYSGSKRHPTK--EKAQPLEDLAGW--KELFOTP--- 1477
Qy 1329 GDVIMEDFNERNSSFTKSHLLSSDAEGNYRDSLETLPSTKESDSYOTTPPSASCESNS 1388
Db 1478 ---VCTDKPTTEKTTKICRCSQPD-----PVDLPSSKRSKR 1513
Qy 1389 VNQVEDMELETSEVKVSS-----PITSEESNLSDNFDIBENGJLPINK--NEWV 1436
Db 1514 SLRKVDVEEFELKRRPRPSACKAMHTPKPAVSGEKNIYA--FM--GTPVOKLIDJENL 1568
Qy 1437 NGESEKRYTIVTMTSTVA-----TESKVIKVEKGDQKVIVSTECARS 1484
Db 1569 TESKRRLTOPKKAQALDELAFKELFOTRGHTEESM--NDKTAQVACKSSQPDLDKN 1625
Qy 1485 TWTTTTTVTKLTSTPGSDVDIISVKEOSKTVTTTVDLSLTGGTLVTSMTVSKES 1544
Db 1626 PASSRRLKTSIG--KVQKEELLAVGKLTQTSGETTHHTHTPTGKGKMAKPAESPKOI 1683
Qy 1545 TPDKVKLMKFSPPKTRSGTA-----LPSTYRK--FYTKSTKKSIFVLLPND 1588
Db 1684 LPSAASLNGSKQLRTPKKSSEVPEDLAGFIELFOTPSHTEKSMTEKTKVSYRASQPD 1743
Qy 1589 L-----KILARKGIREVYFNNAKPAIDIMYPRPRPFGLTMYKRLQTVAS 1637
Db 1744 LVDTPTSSKPPKRSRLRADTEE-----EFIAFRKQTPSAG--KAMHTPKP 1787
Qy 1638 LAGVSLMLR-LIMASLRMDMAKAYPPGGSTRTSETSEITTELIRKRDVGPYGR-- 1694
Db 1788 ANGEKRDINTFLGTVPYOKIDOPGNLP--GSNRLQTRKKKAQALELT-----GFRLL 1838
Qy 1695 -----FEYCIRKIIIC--PIGVP--ETPKETPTPQKGLRSS-----ALNPKRP 1733
Db 1839 FOTPCDTNPTADEKTTKILCKSPSDPADTPTNTRKQPRKSLKADVEEELAFKRLP 1898
Qy 1734 -----EPPKQGVPIIETWVAEEELMEIRAF--AERE-----KE 1768
Db 1899 SAKGAMHTPKAA-----VGEER--DINTFGTPEVKELDLGNLPGSKRRPOTPRE 1946
Qy 1769 KAAAVEEOQAKKRLDEOQKPTVIATSTSTPTSTISPAQKTMVAPISGSVTTGKMYLT 1828
Db 1947 KKALEDLAGFELFOTRGHTEESMTDOKITEVSCSKSPDPVKTP-----TSSQRLK 2000
Qy 1829 TKVGSBAIVTFOQKNFHOTFAWKKOGOSNGVVOQOQVL--GIIPSSTGTSOOTPTS 1886
Db 2001 ISLGR-----GVKKEEVLVPVKGLQTSKTIQT-- 2028
Qy 1887 PQPRATVYIRNTSGSGT-----TSNSQVI-----TGPOIRGMYIYIRP-----L 1929
Db 2029 -----HRETAGDGKSIKAFKESAKOMLDPANYGGMERWP-----RTPKEBAOSL 2073
Qy 1930 QOSTLGKAIIRTPVWQPAQOQVMTQIIRGOVSTAVSAPMTVSSTPGOKLTSATSTS 1989
Db 2074 EDLAGFELFOTRGHTEESMTDOKITK-----ACKSPPEESMDPTST- 2117
Qy 1990 NIOSASOPRPPGOGQVILT--MAQLTDLTQGHGNGOGLTVVIOGOGOTTGOLDLIP--- 2044
Db 2118 -----RRRPKRLPKRDIIVEELSAKOLTO-----THTDKVPGBE 2153
Qy 2045 -OGVTVLGRGQGLM-----QAAMPNGVQ-----RFLF--TPLATATATMS 2083
Db 2154 DGIWIFRETAOKIDLPASVYSGSKRRPRTPKGAORLEDLAGLELFTQTPCTOKPTTH 2213
Qy 2084 TTTTIVSTTA-----AGTGE--QROSKLS--POMOVHODKTLRPAOSSSVGPKAOPOTA- 2134
Db 2214 EKTTRIACRSPQDPVGVPTIFPKRPSKRSRLKADVEEELAKRKTPTSGKAMDTPKPAIG 2273
Qy 2135 -----OPSARPOQTOPQSPAPQEVOTQEVOTQTVSS-----HVP 2171
Db 2274 GDEKDMAKFMGTQVOKLPLPGMLPGSKRWPQTPKE-----KAQLEDLAGFELFQCP 2326
Qy 2172 SEAOPTHAOSSKPOVAAOSQPOSNVQOSPRVAVQSPSOTRIRPS----- 2215
Db 2327 GTDKFT--TDEKTTKIAKS-PO-----PDPVDTPASTKQPRKRNLRKADVEEELALR 2377
```

```
Qy 2216 --TPS-----OLSPGQOSOVOTTTTSOPIPIOPHSL-OIP-SQGQOSQP----- 2256
Db 2378 KRTPSAGKAMDTPKPAVSEKNIINFETRPVOKLIDLGLMLPGSKRQRPQPKKAEALDEL 2437
Qy 2257 -----QVQSTQTLSSQTLNOSVSSPSRQOLQOPQOVLAVP-----QLQOQOV 2305
Db 2438 VGFKELFOTRGHTEESMTDOKITEVSCSKSPOPESEFTSSRSQORLKLPLVKVMKEEPLA 2497
Qy 2306 LSGI--QSQVVAOIQAOQSGVPOQIKL--QLPQIOQSSAVOTHOIQNVV--VOAASVOE 2360
Db 2498 VSKLRTTSETTQTHTEPIGDSKSIKAFKESPRQIIDLPAASVYSGSKROLRTRKRAALE 2557
Qy 2361 QLOQVQOL-----RDQOQK-----KKOQOIEIKR-- 2384
Db 2558 DLVDFKELFSAGHTEESMTIDKNTKIPCKSPPELTDTATSTKRCPTKPRKVEEELS 2617
Qy 2385 --EHTLOASNOSEILIOQVVMHNAVIEHLKO--KKSMTPAEENOMIVCNOVKYTL 2440
Db 2618 AVERLTQTSQSTHTHKEPASGDEG--IKVLKORAKKPPVVEEESRR----- 2664
Qy 2441 DKIDKEKAAKKRRKEEVEQKRSQONATKLSALLFKHEQLRA-----EI 2487
Db 2665 -----PRAPKERAQPLEDLAGFELTSETSGHTQESLTAGKATKIPCESPPELV 2713
Qy 2488 LKKRALDLQOLEVOEELKRLKIKKEKDLQOLAQ----- 2523
Db 2714 VDTTASTKRLHTRQK-----VOYKEEPSAVKFTQTSGETTDADKAPAGEDKGIKALKE 2768
Qy 2524 -ATVAAPCPVTPVLPADPAPPSPP-----PPGVQHTGLSTPTPLVASOKRRR 2574
Db 2769 SAKQTPAPASVYSGSKRRRPARESQAILEDLAGFKDPAAGHTEESMTDOKITKIPCKSS 2828
Qy 2575 EEKQSSSKSKK-----KMIQTSKTEKKDKTLVYICTPTPDESK 2615
Db 2829 PELEDATISSKRRPRTRAQVEKKEELLAVGKLTQTSGETTHFD-----KEPVGEK 2880

RESULT 4
US-09-724-676-65941
; Sequence 65941, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 65941
; LENGTH: 2896
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-65941

Query Match 2.7%; Score 392; DB 5; Length 2896;
Best Local Similarity 18.5%; Pred. NO. 0.0073;
Matches 533; Conservative 370; Mismatches 1025; Indels 948; Gaps 133;

Qy 291 QCEVCVAHKVPVTCVAEIQKNKPYIRH-----EPTG-----YDR 326
Db 34 ECDIRQLRPPVSKOKIEIHQDEAILHNFSSNTPTQVNGSVIDEPVRLKHGDVITIIDR 93
Qy 327 SRKRWFLNRLRIEDPTNENEMKTIWYSTVQVLAELIDCDKQWMEELCKILEMNE 386
Db 94 SFR-----YENESLQNGRKSTEEPRKIR-----DQEPARRVS--RS 127
Qy 387 EIHHRMDITEDLTNRKAGSNKSFLLAANEILE-----SIRAKGDIDN-----VKSPE 435
Db 128 SFSQDPDESEGIPLAKR--RVSGGHLRLELDENLPRPTPLAKREGAPTKRKSLYMHNP 185
Qy 436 ETEKKNETENDSKAEKNRE--EFEDQSL-----EKUSDQKTP--DDPEQKSEVGFK 487
```


Db 186 VLKRIIEQBP6SGKQESSEIHEVEKAQSLVISPAPSPKTPVASDQRRSCKTAPAS 245
 QY 488 SEKSNGLSPGAGCAGASTRIITRLRPDS-----KLSQLKSQVAAAHKANLE 541
 Db 246 SSKSQTEVPR-----RGERVATCLOKRVSISRSQHDLOMICKSRKSGASEAN--- 294
 QY 542 KEQKEVIVVNSQGISRLSTK---EYIMKG---NINNYFKLQEGEGRV--YHNOYSTN 593
 Db 295 -----LIVASMAVYKLGAKQOTKTKHGPORSAMKRRRATPKKPYGEVHSQFSTG 349
 QY 594 -----SFALNKHQHRDHDKR--HLAHPFLTPAGEFK-----WNGSVHSGVL 636
 Db 350 HANSPCTIIGKATEKTEVHPAPRYVLNFIISNOKMDFEDLSGLAEMKTEKTPVEKPOL 409
 QY 637 TISLRLITOLENNIPSS-----LHPMASHRAMIKAYOMCSK--PREPALALA 686
 Db 410 T-STCHTAISSENLKQFOGTDGSEBLLPTSESFGVFFSAQNAKOPSD----- 462
 QY 687 ILECAVPVMLPIWREFLHT-----RLHMTSIERE----- 719
 Db 463 --KCSASP-----PLRQCIENGNAVAKTPNTYKMTSLEKTSDETETEPKSTVTVNRSG 516
 QY 720 ---EKEVYK--KEKQEEBETM-----QOATWKTTPYKHQVWKQKEEYVTC 765
 Db 517 RSTEFRIIOKLPIVSKSEETTELVECILRKQKATLLOQ-----BREGEMKEI-- 565
 QY 766 YGSMWISKTHVYFVKPLQNTVNYRKSLSEGTKNMNDENMDESDK-----RKCS--RS 818
 Db 566 -----ERPET---YKENIELKEN--DEKMKAMKSRWQKOCACAPMSD 603
 QY 819 PKKIRIEPDE--KDEYK-----SDAAGADONEMDISKITEKKDOVYKELLSDSDK 870
 Db 604 LTDLSLPDTFLMOTFARGONLLOTHAKAPKSEKGIITKM----- 645
 QY 871 PKCE--EPHEVDDMKTSHVNCQSSQVADVNVSE-----GFHLRTS----- 911
 Db 646 PCOSILOP---EPINTPLTHKQOLKASIGKVGKEELLAVGKTRTGETTHHREBAGD 701
 QY 912 -----YKKTTR-----SSKLGLERRIKOFTLEEKORLEKIKLESIGKIGTSTNSS 960
 Db 702 GKSTIRTEFSPKQILDPAAARTGM--KKWPTPKREAQSLLEDL--AGFKRLPQTPQPS 756
 QY 961 KNLSEPVITAKBGCOSDKROBOSPANNADOPE-DLIOGCSQSDSVLRMSDPSTHTN 1019
 Db 757 ESMDEKTE---TKIACKSPPEVDPTPTKQMPKRSILRKADVEEFLAKRLTPSAGKA 813
 QY 1020 KLYPKDRVLDOVSI-----RSPETKCPKONSTEN-----DI 1050
 Db 814 MLTPRPAGGDEKDIKAEMGTVPYQKLDLAGLPGSKROLQTPKEKAQALEDLAGEKELFOT 873
 QY 1051 EEKVSDLASRGQ-----EPTKSKTRGNDFFIDSKLASADDIGTLI--CKN 1094
 Db 874 PGHHEELVIAAKGTTKIPCDSPQSDPVPDTSTKOR-----PKRSIKRADVEGELLACRN 927
 QY 1095 KKPLIOESDITIVSSKSAHLSVPKSTNDPD-----ATPLSRAMDEPGKLGCSSESNT 1149
 Db 928 LMP-----SAGKAMHTPKPSVQEEKDIIIFVGTPOKL-----DL 962
 QY 1150 LENSADTVISIDSSPEEDMIVONSNMESISSEOPRTREDOVEVLEPLKELVSGESTG--NCE 1208
 Db 963 TENLTGSKRRQOTPKREKQALEDLTGFKELFOTPGHTEEA-----VAAKTKMKCES 1015
 QY 1209 RLVPVGTBANGKPSQOKLEERPVNKCSDQIKLNTDKKNNRESEKKGGQSTFQOI 1268
 Db 1016 SPPEBADOPTSTBRQPKPLERKDYOKELSAK-----KLTQTSGETTHHDTK 1063
 QY 1269 NGKDNKPKIYIKGECIAKISSRVYSGNV--EPKYANINIKIIPENDIKSLTYKESAIRPFI 1327
 Db 1064 PGGEKKSINAPRETAOKOLDPAASVYTGSKRHPKTK--EKAOPLEDLAGW--KELEFQRP-- 1117
 QY 1328 NGDVIMEDEFNENSETSKSLSSSDAGNRSDETLPSTKESDSQUTTPPASCPESN 1387
 Db 1118 ---VCTTKPPTTHEKTTIACKRSQPD-----PVDIPTPSKPSQK 1152

QY 1388 SYNQVDEMELETSEVKYKTS-----PITSEESNLSNDFIDENGLPINK---NEN 1435
 Db 1153 RSLRKADVEEEPFALLKRRPPSAGKAMHTPKPAVSGSKNIYA--FM---GTPOKLDLTEN 1207
 QY 1436 VNGESKRKTVITEVMTSTVA-----TESKVIIVVEKQDKQTVYSTENCAK 1483
 Db 1208 LTGSKRRLQTPKREKAQALEDLAGEKELFOTRGHTEESMT---NDKTRAKVACKSOPDLK 1264
 QY 1484 STYTTTTVTKLSTSTGSDVILSVKQOSKVTYTTVTDLSITTTGTLTSMYSKEX 1543
 Db 1265 NPASSKRRLKTSLG--KVQKEELLAVGKLITQSGETTHHEPTPGDGSMAPFESPOQ 1322
 QY 1544 STRDKVKLMEKFSRPRKTRSGTA-----LPSYRK--EYTKSTKKSIFYLPND 1587
 Db 1323 ILDSASLGSKQRLTPRGKSEVPEDLAGFIELFOTPSHTKESMTNEKTKTVYSKASQ 1382
 QY 1588 DL-----KILARKGITEVYFYNNKAPRALDIMPSPRPTFGITMYRLQTVK 1636
 Db 1383 DLVDTPTSSKPOKRLRKADTEE-----EPLAFRKQTPSAG---KAMHTK 1426
 QY 1637 SLAVSILMR-LIMASLRMDMAKVPYPPGSGSTRTESETEITTEIIRRDVGPYIR- 1694
 Db 1427 PAVGEKQDITFLGTPOKLDOPGNLP--GSNRLQTRREKAQALEELT-----GFRE 1477
 QY 1695 -----FEYCIKTIIC--PIGVP--EPKETPTPOKGLRSS-----ALPBRK 1732
 Db 1478 LFQTPCTDNPTEADEKTKTKILCKSPQSDPADPPTNTKORPKRSILKADVEEFLARKLT 1537
 QY 1733 P-----EPKQGPVITIEVWAVEELELMEIAF-----AEVRE-----K 1767
 Db 1538 PSAGKAMHTPKAA-----VGEK--DITFVGTPEYKLDLGNLPGSKRRQOTK 1585
 QY 1768 EKAQAVEOQAKRLEQOKPTVIATSTTSPTSSTTSISPAQKVMAPVIGSVTTGKRWL 1827
 Db 1586 EKAKALEDLAGEKELFQTPGHTESMTDKITFVCSKSPQDPVKNP-----TSKQRL 1639
 QY 1828 TTVYVSPATVTPQONKNFHOTFAVWKQOGSNGVYQVQOKVL--GLISSQTSQOT 1885
 Db 1640 KISLKG-----GVKKEEVLVYKLTQTSKQTTQ-- 1668
 QY 1886 SPQPRATVTRIPNTSGSGT-----TSNSQVI-----TGPOIRGMVYIRTP----- 1928
 Db 1669 -----HETADGSGISAKFESAKQMLDPANYGIMGRWP-----KRPKEAOS 1712
 QY 1929 LQOSTLEKAIIRTPVWQVPAPOVMQITIRGQVSTAVSAPNTVSSSTPGOKSLTAST 1988
 Db 1713 LEDLAGKELFQTPDHTEESTDDKTKI-----ACKSPPEESMDPTST 1757
 QY 1989 SNIOSSASOPPRQOGVOKLT--MAQLTOLTQGHGNGGLTVYIQOGGOTTGQLILP-- 2044
 Db 1758 -----RRRPKTPPLGRDIVEELSAKLQLO-----THHTDKVPGD 1792
 QY 2045 --OGVTVLPBGQOLM-----QAAMPNGTVO-----BFLF--TPLATTATTA 2082
 Db 1793 EDKGINVFRETAKOKLDPASVYTGSKRQPTPRGKAPLEDLAGLKELFOTPVCTDKPT 1852
 QY 2083 STTTTVSTTA-----AGTGE--QROSKLS--POMQVHQDCTLPPAOSSSVGPAAKAPOTA 2134
 Db 1853 HEKTTK IACRSPODPVGTPTIEFKPOSRLRKADVEEESLALRKRTPSGKAMDTPKPA 1912
 QY 2135 -----QPSARPOPOPOPSAPOEVOGTQPEVOGTQTVSS-----HV 2170
 Db 1913 GGDEKDKAMFGTPVQKLDLPGNLPGSKRWPOTPE-----KAQLEDLAGEKELFOT 1965
 QY 2171 PSEAOPHTNAOSSKPOVAOQOSPNVQOSPVVYOSPSQTRIRPS-----KAQLEDLAGEKELFOT 2215
 Db 1966 PGTDKPT--TDEKTTKICKS--PQ-----PDPVDIPASTKQKPRKRLRKADVEEFLAL 2016
 QY 2216 ---TPS-----QISPGQOSQOVQTTTSQPIPIOPHTSL--QIP--SQGQPOQOP----- 2256
 Db 2017 RKRTPSAGKAMDTPKPAVSEKININTEVETVPQKLDLGNLPGSKRQOPQKREKALEBD 2076

QY 2257 -----QVOSTOTLSSGOTLNOVSVPSPRQLOIOTQPOPOVAVP-----OLOOQVQ 2304
Db 2077 LVGFELEFPTPGHTEESMTDDKITEVESCKSPQESFETSRSSKORIKITIVKMMKEEPL 2136
QY 2305 VLSOI--OSQVAAQAOQAGSVPOQIKL--OLPIQOQSSAVQTHQIQWVVT--VQASVQ 2359
Db 2137 AVSKLRTISGETTQHTTEPTGDSKSIKAKESKQIOLDPRAVSYGSRRDLRTKREKARAL 2196
QY 2360 EQLORVQOL-----RDOOK-----KKOOOTEIKR-- 2384
Db 2197 EDLVPEKELFSAPGHTEESMTIDKNKIKPCSPPELDTATSTKCRPKTRPEKEEEL 2256
QY 2385 ---EHTLQASNOSEIIOQOVYKHNVAIEHLQ--KSMTPAREENQMIYONVYKTI 2439
Db 2257 SAVERLTQTSQSTHMKRPPASGDEG--IKVLKORAKKPNPVEEESRR----- 2304
QY 2440 LDKIDKEEQAKKRRKRESEVQKRSQONATKLSALLFKHKDOLRA-----E 2486
Db 2305 -----RPRAPKKAQPLEDLAGFTLSETSGHTQESLTAQKATKIPCESPLE 2352
QY 2487 ILKRALDKDQIEVQELKRDLIKKEKDLMOLOA----- 2523
Db 2353 VYDTASTKRHLRTRYQK-----VQKEEPSAVKFTQTSGETTDADKEPAGEDKGIKALK 2407
QY 2524 --ATAVAACPPVTPVLPAPRPPSP-----PPPGVQHTGLISTPLTPVASQKKR 2573
Db 2408 ESAKOTPAAPASVTSRRPRAPRESQAIEDLAGFKDPAAGTHEESMTDDKTTKIPCKS 2467
QY 2574 REEEKSSSKSKK-----KMISTSKETKDKDKLVCICKCTKPYDESK 2615
Db 2468 SPELEDTATSSKRRPRTRAKVEVKEELLAVGKLTOTSGETHHTD-----KEPVGEK 2520

RESULT 5
US-09-724-676-69138
; Sequence 69138, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69138
; LENGTH: 4168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-69138

Query Match 2.7%; Score 390.5; DA 5; Length 4168;
Best Local Similarity 16.8%; Pred. No. 0.011;
Matches 583; Conservative 479; Mismatches 1180; Indels 1223; Gaps 130;

QY 17 QDSDEDEDEMEDDDSDSPREMEDDDDASTCESRSSTSYSTGRRKPRVHRP 76
Db 1040 KTESDQDDEIKTRKQSF-----ASIALRKRYSLTEPGMSEQEURSK 1085
QY 77 SP-----ILEKDIPL----- 89
Db 1086 IRPHLLLOALLFYKTIPLVLDSSDYSAMASOYRLHLFIKFLIYLAISVYKIMVCFPT 1145
QY 90 -----FKSSEDLVMPNEHIM--NYIAI-----YEVLN 116
Db 1146 FSNQIHIRRVNTIFSKSISDVASPIQILTDNVFADKNCGVTSIQYPGFLWPGYSSSH 1205
QY 117 FCTVRLSPFRFEDCALVSOQCTLAEMHNVLIKAVLREEDJTSNTFG-----P 168
Db 1206 GSYALKRAGIQYVFLSNLS-----DYSRVSGEYINMTTP 1243
QY 169 ADLKDSVNTLFLFIGMTPVELRVYCESDKYHNHLYQ-----EADDPYGVENKIKIV 224
Db 1244 ASPKSNIM-----YSSSLPKRSITTSAPLISPLKSVSP 1280

QY 225 LQFLVDQFLITNIAREELMSEGIYQDHCRCYKILGL-----LCETCSAVYH 275
Db 1281 VKSRVDIVISAKITMASLSSIPVKQMPGHAVALVNGSISPLKAYSSSTLINCKRATATL 1340
QY 276 ECVKPRELEVEDMOCCEVCA-----HKVGVTCVCAEIQKNKTYIHRPEIGYRSR 329
Db 1341 Q-----EKISSATNSVSSVSAADTVKEVFTSTDCNA--IFPTQVLCFSTISFSVSKN 1393
QY 330 KY-----WFLN--RRLIEEDTENENBKIMYSTVOAELIDCDKDVWEAELK 379
Db 1394 SFRKCTIYIPWVNYICNYLICNFININSARYSVYVNLPEPALKLDPDSFTRKSAALLS 1453
QY 380 ILEEMREELH--RHMDITDLTKRAGSNKSFIA-----ANNEELIESIRAK 425
Db 1454 PIKTLFTETHPOHPFRTSSPV-----KSLFLAPALKLSPSSLSSEOELIKDYAEWK 1508
QY 426 GQIDNVK-----PEETKDKNETENDSKDKENREEDDS--LEKQ----- 466
Db 1509 EDLMRTAILQTDVPEKPFQPELPKEGRIDDEEPKIVEKEDLVKYSILTKQVCVD 1568
QY 467 -----SDDK--TPDDPEQGSSEVGFSEKSGELSESPG-----AGKAS-- 506
Db 1569 NGSPKSPKSDKGHPEDWIERSE--EIRARQQAASQSPSLPERVQVAKAASEKD 1626
QY 507 -GSTRIITRLRNP--DSKLSQK-----SQOV--AAAHANKLFKEGSEVL 548
Db 1627 YNLTKVIDYLTNDIGSSSLNLKYPEDAKKDQEGQKRVLKALQEHKLMPPASMR 1686
QY 549 VANSQGEIRLSRK---KEVIMKGINNNFKLQGEKRYVYHNQYTSNFALNKHQRED 605
Db 1687 TSTSEKELCKMADSPFGTJILES--PDGFQHDQKSPLSQSGFETREKTPSAPQSAE 1744
QY 606 HDKRRILAKFCFLTPAGEFKMNGSVHGSVLTISTRLTTOLENNI--PS-SFL-- 658
Db 1745 TTGPKFLFHEVPLPPVITETREYVAVHISYDPSADVDQOTQPEEYVSRKPSFTHELP 1804
QY 659 -PNMASHRANWIKAYOMCSKPREFALATALDCAVAPVYMLPWREFLGHTRLHRMTSIE 717
Db 1805 KPTTSIKERF--KAFQM-----KASSE 1825
QY 718 REKEKVKKKKKEKQEEELMQATVYKTPRPYKHOVWKQGEERYVYTGSMWSIKTYV 777
Db 1826 EDDHNRVLKGMKVKDETHITTTTRVYHSP----- 1856
QY 778 YRFVPLPGNTNVNRYKSLGTRKNMDEMDSKRCRSRK-----KIKIE 825
Db 1857 -----PGG-----EGASERIEFTMSYHDIKAKAQSGHDPSEKELAGLEFKRSAYS 1900
QY 826 PD-----SEKD-----EV--KGSDA-----AKGADQNEMDI- 849
Db 1901 PVDHKAALFSAQHAOKDQNMKPLERIIIEVHEKQNAKEPTEVITRETKHREKEMYY 1960
QY 850 SKITTEKKDDYKELDSDSK--PCKEPEVVDMDMT--ESHVNCQESSQVNVVYSEGF 906
Db 1961 QKDLISGDIINLKDPLPEKIDAPFSEQOQOEEELTJAEBSLPSYLESSRYV--TPVSOE 2019
QY 907 HLRTSYKK-----KTKSKL-----DGLERIRIOFTLEEKORL--EKIKL----- 945
Db 2020 DSRPSSAQLISDSYTKLLLSLHSHSIEYHDELSELRGSGSYRAEAMLLSEKLDVSHST 2079
QY 946 -----EGIKIGIKSTNSKSLSESPVITKAEQGSQSDSMROQSPNANDOP 994
Db 2080 EESVTHAGPPESELOGSDKRSREKJATAPKKEIISKIKYDVSENGVGK-----VSKDEH 2134
QY 995 EDLIQCSQSDSSVLRAMPSPHT-----TNKLYPKRDYLDVYSIRSPETKCP 1041
Db 2135 FDKVTVLAYSQGN---YSSPKHAMMRFTEDRLRGREKLIYEDRV--DRTVKEAEKELT 2188
QY 1042 KONSIENTDIEEKVSDLASGOEP--TKSKTKGNDFFLDDSKLASAD--IGTLICKNKK- 1096
Db 2189 EVSOFPKDTKEKLDN--ELQSEPKARPKNGKEYSSQSPSTSSPKVLLTTELLASNDWM 2245


```

QY 1097 -----PLIOESDITVSSSKSALHSSVPKSTINDRAT-----PLSRAMD- 1135
Db 2246 VKAROHGPDGQEPKAEKPSLSPSSPEKKVL- -QOTEDSKSTVEAKSISQSKAPDG 2302
QY 1136 -----FPGKLGCOSESNSLTLENSSDTVS-----IIDS- 1162
Db 2303 POGSFOLOKOSKLSIRLKEFOGTHAKSKDMSQEDRKSDDGOSRIPVKKIQESKLPVYOFA 2362
QY 1163 -SEEDIMVONSNEISIO-----FTRQDQVEYLE----- 1191
Db 2363 REHQOKAIDLPEDESYSQOKFVWLKTRDEHAQNSNEIYVNDGSDNVKQKOTEMSKAMPD 2422
QY 1192 -----PLKCELVGSESTGNCEDRLPVKGTANG--KKPSQOKLIEEPVKKCSQDK 1241
Db 2423 SFSEQOAKDLACHITSDIARFGPMDKKVFRTWESSGATNNKSQEKLSHVLH----- 2475
QY 1242 LKNTTOKKNE--NRESEKKGORTSTFOJNGKNNKPRIYKGCGLKISSRVAGNVER 1299
Db 2476 -----DVRENHIGPEKSKSYQOKNEKMSYTEREK--LITNGSLSEIKEMTVKSPS-- 2524
QY 1300 KVNINKIIPENDIKSLTVESAIRPFIINGDVIMEDFENRNSSEKSH-----L 1349
Db 2525 -----KKVLYREYVKBGDHPGGLDQPSRSESSAYSIIHVRVADERML 2570
QY 1350 SSDBAGNVDLSLETLPTKESDSTQTTTPSASCPSNSVNOVEDMEIETSEVAKVT-- 1406
Db 2571 SSNIPDGFCEOS--AFPKHLSQKLSOSSMSKETFVETQHNLSIEDEKVYSEISKVSKHQ 2628
QY 1407 -----SSPTSEE-----ESNLNDFIENGRLPKNNENV-----NGES 1440
Db 2629 SYVGLCPBLEETSTSPKSPDSLEFSPGKSPSSDVFDS--PIDGLEKAPLAQTEGK 2686
QY 1441 KRRTVITEVTMTSTVATESKTVIKVEKGDQTVSVSTENCAKSTVTTTTVTKLSTPS 1500
Db 2687 EITLPLVYV--SPVOYGQKYEKEIQGGVKKIIISOECTVQETRTFTTTQOQKPPS 2742
QY 1501 TGSVDIISVKEOSKTVYTTTVDLSITTTGGLTVLSMT-----VSEKYSRDKVKLMMK- 1554
Db 2743 POGSPEDDTLEQVS-----FLDSSGKSPLEETPEPSESEVSEYFTKSTDSLAVI 2792
QY 1555 -----SRPKTRSGTALPSYRKFFVTKSTKSLFV--LPNDOLKLARGGI 1598
Db 2793 PGKRPSPREVSSESEBEQAKSTSL-----KQTVVEETAVREMBNDVSKDSNOGPKN 2845
QY 1599 REVYFENYNAKPAID----- 1618
Db 2846 NRVAYIEFPRPLDADQIESDKKHNHLYPEKEVDMIEVNLQDEHDKYQLAEPIVRQP-P 2904
QY 1619 SP-----RPTGITWRYRLQTV-----KSLAGVSLMLRLMMS-- 1651
Db 2905 SPVPPGADVSDSSDDESITQVPVKRYTFPLKEVDEDEQEKPRKASAEKASNOKELESNGS 2964
QY 1652 -----LRDDMAKAVPPGSGSTRSTSETTEITTEIIRRD-----VGPGRIRE- 1696
Db 2965 GKDNREGGLGIDSPONELAQNGNNDOSTTECSITTAFFSHDTATLETDSLDGDLQDEDD 3024
QY 1697 -----YCIKRIICPIG-----VBETPKETPTPOKGLIRSS 1726
Db 3025 GLTESDSKLPQAMEIKKIDIMWTEGILKPADBSFQSOSKLEVEIEEGKVGPDDEKPPSKSS 3084
QY 1727 ALRPKRPETPKQTVPIETWVAEELIEMETAFARVEKEKAQAVEQOA----- 1777
Db 3085 SSEKTPDKIDOKSGAOFTELEGHNPDRSVFPDYFFSKYDEEFATPFKIVATGLDEDPW 3144
QY 1778 -----KKRLQOKPVIATSTTSPSTSTSTSPAKVAVAPISGVTTGTGM 1825
Db 3145 SNNRGDEVDSKSRDEETKPEGLAVEDRSP-ATTPDTPPAR-----TPP 3188
QY 1826 VLTTKVGPATVTFQONKNHOFATMV-----KOGS 1858
Db 3189 DESTPTSEPRPFHFGKMEEMTRSGAIDMSKRDYERLQFOIGEHNSEKSGOQGES 3248
QY 1859 NSGVVVOQKVLGIIIPSSGTGSOQFTTSFOPRTATVTIRPN-----TSGSGGT 1906

```

```

Db 3249 DKSMVATAPQ-----POSGDTVE--TNLERNVETPVEPNPISPTSGEQEGSTSSGSL 3301
QY 1907 TSNQSVITGQPIRPGMIVITPPLQOASTLGKAILRTPIYVQPGAPQOVMQI-----IRG 1960
Db 3302 EKSAATNTSKVP--KLNTPIKM-----GISASTYMKKEPPELTDTKIEVAVTSCOG 3353
QY 1961 QPVSTAVSAPNTVSSSTPG-----OKSLTSAT-STSNIOS-----SAS 1996
Db 3354 LEMETITMISNTANSQGVAPRHEKHDPQKDNFNNNNNLDSSTIOTDNIMSNIVLTHEAP 3413
QY 1997 QPPRPQOGVYKLMALQTLQTOGHG-----GNGLYVYIOGQO----- 2035
Db 3414 TCTTKDNVPKVSCKTGVLOGCVARDKQVLEQOQKTELIGIROKSKLPKATSPKD 3473
QY 2036 -----TTGQLDLPQGVVLPFGPQGLMOA----- 2060
Db 3474 TFPNHSNTKASKMKQVOSSEKTKALITSSCYDVKSRIIPVKNTPRDNITIAVKKACATQK 3533
QY 2061 -----AMPNGTVORFLPTPLATVATTASTTTVYSTTAAG----- 2095
Db 3534 QGQPEKGAQQLPSKLPVKKVSTCVTTTATTTTTTTTTTTTSCVVKRSQLEKVCQ 3593
QY 2096 -----TGE-----QROKLSPOMYHODKTLP-----PA----- 2119
Db 3594 HSIEYKGISGETLKLVDRLSEBEKKMOSELDEEEKYLKHVVVRDPFGWPAFGYKKS 3653
QY 2120 --OSSVGPAAKQAPQTAQPSARPOQOTOPQSPAPQEPQVQTPREVQTOQTVSSHHPSQPT 2177
Db 3654 RDKKTEAPLKSSEKAGSKSRSSRRTPQSPCE-----RDIIMAIIVAD----- 3698
QY 2178 HAQSKRVOVAQOSPOSNVQOSPVVROVSPQO----- 2209
Db 3699 HLGISMTLAREL--NFSVDEINQIRENPNLSIOSFMLKKWVRDQGNATDALTGV 3756
QY 2210 -TRIRP-----STPSQLSPGOQSOVQYTT-----SQ 2234
Db 3757 LTKINRIDIVTLLEGPITFDYGNISGTRSPADENNVRHDPVGPSLOVELEPTGLHYTP 3816
QY 2235 PIPLOPH-----TSLQIP-----SOGPOSOPOVQOSTQTLSSGQTLNGVSSPSR 2281
Db 3817 PTFPOQDDTFSDLSSTIESPLRTPSRLSDGLVPSQGNIEHS--AGCPVYVTAEDASLED 3872
QY 2282 POLQIOOP--QPOVIAVPOLQOQOVOLISQIOSOVVAQIOAQSQVPPQIK-----LQDPI 2334
Db 3873 SKLEDSVPLTEMEADVDESQLENVCLSE-----YPOYLGNLAGSPKDVKPAEPKRLGV 3927
QY 2335 QIOSSAVQTHQIQNVYVOAASVQEOLORVQOLRQOQOKKKQOQIEIKREHTLQASNS 2394
Db 3928 SSEQOEGKSGSPDEEMMEELKSLFEDIOLEEGVESEEMTEEKVQALIKR-----VQOA 3981
QY 2395 EITOKOVVMKHNAV-----IEHLKOKKSMTPA--EREENQMIYCNQVMYIILDKIDKE- 2447
Db 3982 ELEMSTITGQNTSSGNLSESCAQARVYIGGLDRIDDSQDQORDSITSTYKGEAKFPA 4041
QY 2448 -----KQAAKRRKREESVEQ--KRSQONATKLSALLFKHEQOLRAELTKRALLDKD 2497
Db 4042 NGSHTLETPEAKTKSYFESQNDVQKOSTKETLKPRIHSGHVEBPASPLAAYOKSLEET 4101
QY 2498 LQIEVQEBLKRLKIKKENDLQLOAQATAVAAFCPPVTVYLPAPRPAPSPSPRPVQVPT 2557
Db 4102 SKLIDEE-----TRKCVVSMKMSKRSIPADGKP----- 4130
QY 2558 GLISTPTLPVASQKRRKEEKSSSKKKKKMISTSEKKTDTK 2602
Db 4131 -----RLSLHEEGSGSGEQKQOGBGKV--TKKEIR 4160

```

RESULT 6
 US-09-724-676-69127
 ; Sequence 69127, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO: 69127
LENGTH: 4189
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-69127

Query Match
Best Local Similarity 16.7%; Pred. No. 0.011;
Matches 580; Conservative 474; Mismatches 1179; Indels 1241; Gaps 129;

QY 17 ODSDEDEDEDEDDDDSYREMEDEDDDDASCTESSFRSHSTSYSTGRKRRPRHRR 76
DB 1061 KETESDODELEKTRDROSF-----ASLALKRKYSLTEFGMESEQERYSMK 1106
QY 77 SP-----ILEEKDIPLE----- 89
DB 1107 IPRHLLIALLFYKTIPIVLDSNDYSAMASQVRLHLKFLFYAISFVKINMYCPTD 1166
QY 90 -----FPKSEDLMPNENIM-NYAI-----YEVLRN 116
DB 1167 FSNQIHRRVNYIFSKSIDVASPIQDILTDNFAKCGVTISIQPGFLWPGYSSSH 1226
QY 117 PCTVRLSPFRFEDCAALVSOQCTLAEMHVLLKALREDPNTFG-----P 168
DB 1227 GSYALKRGIQYVFLSNLS-----DYSRVFGEYINMYTPP 1264
QY 169 ADLKSVNSTLFIQDMTEPEVLRVYCESDKENHVLRYQ-----EAEDRYGVEKKIKY 224
DB 1265 ASPKSIIN-----YSSSLPKSIITSAAPLISPLKSVGP 1301
QY 225 LQFLYDQFLTYIANEELMSEGYIQDHCRCVCHKGLD-----LCCTCSAVYHL 275
DB 1302 VKSRYDVISSAKITMASSLSISPVKQMGHAEVALVNGSISPLKAYASSSTLINCKATL 1361
QY 276 ECVKRPLEEVREDEMOCCEVCA-----HKVPVTPDCVAEIQKNKRYINHEPGRYSRR 329
DB 1362 Q-----EKISSATNSVSSVSAADTYEAKVSTTDCNA-ITPTQVICCTSTSEFSVSKN 1414
QY 330 KY-----NFLN--RLLIEDTEENENEKKIWTYSTVQVLAELIDCLDKYWEAELCK 379
DB 1415 SRRKCTLYIPWVYVNCYILCNFINVNSARYVNVLPALPKLPDSNFTKSAALLS 1474
QY 380 ILEEMREELH--RHMDTEDLTNKARGSKSFLA-----ANNEILIESIRAKK 425
DB 1475 PKTLTTEHPQHPFSKRTSSPV-----KSLFLAPALKLSTPSSLSOSSOELLKDAEMK 1529
QY 426 GDIDVYKS-----PEETEKDKENTENDSKOAEKREPEEDS--LEKD----- 466
DB 1530 EDLMRTALQTDVBEKRFQELPREKRIIDDEPKIYEKVEDLVKVELIKKQVCVD 1589
QY 467 -----SDDK--TPDDPEQKSEVGFKESEKNGELSESFG-----AGKAS-- 506
DB 1590 NKGSPKSPKSDKSHPEDDWIEFSSB--EIRARQQAASQSPSLPERQVAKAASEKD 1647
QY 507 -GSTAIRLRANP--DSKISQLK-----SQQV--AAAANEAKLFKEGKEYL 548
DB 1648 YMLTIVIDLTDNGISSSLTNLKRYFEDAKKDEGQKRLKRALDQHKLMPPASMR 1707
QY 549 VNSGGEISRLSTK--KEVIMKGINNFKLGQEGKHYVYINQYSTNFALNKHHRED 605
DB 1708 TSTSKEKCKMADSEFGDTILLES--PDDESQHDDKSPLSOSGEFTRSEKTPSAPQSAE 1765
QY 606 HDKRRHLAHKECLTPAGEFKMANGSVHSGVLTITSLRLTLTLENNI--PS-SFL-- 658
DB 1766 TTGPRRLFHEVPIPVITETREVVHVIRSYDPSAGDVQIQTPREEVSKSPSTHELEP 1825
QY 659 -PNWASHRANMIKAVOMCKSPREFALALALECAVAPVVMLEPIWREFLGHTRLHMTSIE 717

DB 1826 KPTTSIKERF-KAFQ-----KASSE 1846
QY 718 REEKEVKKKKKEKEEETMOQATWVKYTPRYKQWQKQGEERYVTGYGMSWIKTHV 777
DB 1847 EDDHNVLSKGRVKEETHITTTTRVYHSP----- 1877
QY 778 YRFVPLPGNTVNVYKRSLEGRKNMDEMDSDKKRCSRPK-----KIKIE 825
DB 1878 -----PGS-----EGASERIEETMSVYDIDKAPQSGDPSKLAGLEPHKSAVS 1921
QY 826 PD-----SEKD-----EV--KSDA-----AKGADQNMEDT- 849
DB 1922 PVDHKSAAETSAQHAENKQMDQMKLERIIEVHLEKNOAPEVILIRETKHPEKEMYYV 1981
QY 850 SKTTEKDDQVKELLDSDSK--PCKEPEHVDQDMT--ESHNCQESSQVNVVNSBEG 906
DB 1982 QKDLSGDITNLKDFLEKHDAPPCSEGOQOEELTAEBSLPSYLESSRVN--TPYSQEE 2040
QY 907 HLRTSYK-----KTKSKL-----DGLERRIQFTLEEKORL-EKIKL----- 945
DB 2041 DSRPSSAQILSDSYTKLLLSQHSIEYHDELSELRGESYRAEKMILLSEKLDVSHDT 2100
QY 946 -----EGIGIGKGTNSKKNLSESPVITAKEGCOSDSMRQSPNANNOP 994
DB 2101 EESVTDHAGPSSSELOGSKDRSKREKATAPKKEILSKIKYDVSENGVGR-----VSKDEH 2155
QY 995 EDLQCSQSDSVLMDPSHT-----TNKLYPRDRLDVYSIRSPFKCP 1041
DB 2156 FDKVTVLHYSN-----VSSKHAMMRFTEDRLDRGREKILYEDRY--DRTVAEAEKLT 2209
QY 1042 KONSIEDEIEKVSIDLASGOEP--TKSKTKGDNFIDDSKLSAAD--IGTLICKNKK- 1096
DB 2210 EYSQFPRDKTEKND--ELQSPKKARKPKNGKEYSSQPTSSPKVLTTELLANDEM 2266
QY 1097 -----PLIOESDITVSSKSALHSSVYKSTNDRDPT-----PLSRAMD- 1135
DB 2267 VKARQHPGQGPRAKEEAPSLPSSPEKRVLS--QOTEDSKSTVEANGSISQSKAPG 2323
QY 1136 -----FEKGLGQSESNSTLENSDYS-----TODS----- 1162
DB 2324 POSFQLOKSKSLSTLKEQGTNAKSKDMSQEDRKSQOSRIPLVKKIOESKLPYQVFA 2383
QY 1163 -SEEDMIYVONSNEISQO-----FTRQEDVEYLE----- 1191
DB 2384 REKQAKAIDLPOESVSVQDFVLTAKTDEHAQSNELIVNDSGDVYKQKOTEMSKAMP 2443
QY 1192 -----PLKCELVSGESTGNCEDRLPYKGTBANG--KKPSQAKLEBRPVYKCSDJT 1241
DB 2444 SFSEQOAKDLACHITDLATRGPMDKKVFRTWESSGATYNNKSQEKLSHVLVH----- 2496
QY 1242 LKNTTKKNNE--NRESEKKGORTSTFOJNGKDNKPKIYLGCELCLEISRRVNGVEP 1299
DB 2497 -----DVRENHGHPSKSVYDQKNEMSVTERERK--LITNGSLSEIKEMTVKSPS-- 2545
QY 1300 KVNINIKIIPENDIKSLTYKESAIRPFINGDYIMEDFERNSESTSH-----TL 1349
DB 2546 -----KKVLYREYVVKGDHPGGGLDDPSRSESASAASHIPVRADERRL 2591
QY 1350 SSSDAEGNTRDSLETLPSRKESDSTQTPPSASCPSNSVNOVEDMEIETSEYKVT-- 1406
DB 2592 SSNIIPDGFEQGS--APRKHELSQLQSSMSKRTVETQHFNSIEDEKLYVSEISKYKHQ 2649
QY 1407 -----SPTTSEE-----ESNLSNDEIDENGLPINKNENV-----NGES 1440
DB 2650 SYVGLCPLEETETSPTKPSPDLSFSPGKESPSVDHDS--PIDLEKLAPLAQTEGSK 2707
QY 1441 KRRTVTEVTMTSTVATSEKTYIVYKEKGDQTVVSTENCAKSTVTTTTVTYTKLSTPS 1500
DB 2708 ELKTLPLVYV-----SPVQVQKQYEKEIQOGGVKKIISQGEKTYOETGTYTTRQOQOPS 2763
QY 1501 TGSVDITISVKQSKTVTTVTYDSTLTGTLVTSMT-----VSKYSTTRDKVKLMKF- 1554


```

Db 2764 PQSGEDDTLEQVS-----FLDSSGSKPLTPEPSESSEVSEYFTSKTPDLSLAVI 2813
QY 1555 -----SRPKRSGTALDPSYKRFYTKTKKSTFV---LPNDLKLARKGCI 1598
Db 2814 PGKSPPIPEVSESEEEQAKSTSL-----KQTVVEETAVEREMENDVSKSDNORPKN 2866
QY 1599 REVPYFNNAKPLLD-----KQTVVEETAVEREMENDVSKSDNORPKN 2866
Db 2867 NRVAATIEFPPLDADQIESDKHHYLPKEVDMIEVNIQDEHDKYQLAEPYIRQV-P 2925
QY 1619 SP-----RPTFGITWRVRLQTV-----KSLAGVSLMLRLMLAS-- 1651
Db 2926 SPVPBGADVSDSDSESYQVPVPVKYTFKLEVDDEQKPKAKASEKASNOKELESNGS 2985
QY 1652 -----LRMDMAKAVPBGSGSTRTETSETTEITTEIIRKP--VGPGRFE- 1696
Db 2986 GKDNERGLGDSQNEIQAONGNNDOSTIECSIATTAEFSDHDTATEIDSLDGLDDEDD 3045
QY 1697 -----YCIKRIIICPIG-----VEPTPEKPTTPQKGLRSS 1726
Db 3046 GLTESDSKLPQAMEIKKIDWNTGILKPADBSFSOSKLEVIEBEGVVGDEDEKPPSKS 3105
QY 1727 ALRPKRPEPRKQTPVLIETVAEELEWEIRAFAREKEKAQAVEQQA----- 1777
Db 3106 SSEKTDKTDQKSGAOFETLEGHHPDRSVFPDITYFSYKVDDEEFATPEFKTATGLDPDPW 3165
QY 1778 -----KRLQKQKPVIAITSTSTSTSTISPAQKVMAPISGSVTTGTM 1825
Db 3166 SNNRGDEVDKSKREDETKPGLAVEEDNSP-ATTPDTTPAR-----TPT 3209
QY 1826 VLTTKVSPATVTFQONKNFHQTFATV-----KOGQS 1858
Db 3210 DESPTPEPRPFPHEKMEFMTRSAIDMSKRDVEERLQFQIGHEHTEGSKGDQGEF 3269
QY 1859 NSGVQVQAVGLIGTSISTSOQTFSFQPRATVYIRN-----TSGSGGT 1906
Db 3270 DKSMVATPQ-----POSGTIVE--TNLERNVETPVEBENPSIPSGECQSTSSGSL 3322
QY 1907 TSNQSVITGQIRPGMIVITPLQOSTLGAIRTPVMVOPGAPQOQMTQI-----IRG 1960
Db 3323 EKSAATNTSKVDP--KLTPPIKM-----GISASTMTMKKEGGETTDIEAVMISCG 3374
QY 1961 QPVSSTAVSADNTVSTPG-----QKSLTSAT-STSNIGS-----SAS 1996
Db 3375 LENETITMISNTANSQMGVAPRHEKHDFQKDNFNNNNNLDSSTIOTDIMSINIVLTHEAP 3434
QY 1997 QPRPQOGQVKLMAQTLQTLQGHG-----GNQGLTVVIOGQCG----- 2035
Db 3435 TCTTEKDNPVKSSGKKTGVLOGHCYARDKOKVLGEQOKTELIGIRQKSKPLIKATISPKD 3494
QY 2036 -----TTGQLQILPQGVIVLPGPGQQLMQA----- 2060
Db 3495 TTFPNHMSNTRKASKMKQVSOSEKTKALTITSSCDVKSRIIVKMTPRNIIIAVKKACTQK 3554
QY 2061 -----AMPNGVORFLFTPLATTAFTASTTTTTVSTTAAG----- 2095
Db 3555 QGQPEKKAQOLPSKLPVKVRSCTVTTTTTATTTTTTTTTTTTTSCHVYKKSQLEK 3614
QY 2096 -----TGE-----OROSKLSQMOYHODKTL-----PA----- 2119
Db 3615 HSTIEYFEGIGETILKLVDRLSSEBKMKOSELSDEEEKYLLKHVVYRPPCPMAFGYKSA 3674
QY 2120 --OSSSVGPAKAPQTPQAPSPAPQTPQSPAPQEPVQTPQEPVQTPQTVSSHVSEAPPT 2177
Db 3675 RDKKTEAPLPSKSEKSGSRRTGPOSPCE-----RNDIMAIAD----- 3719
QY 2178 HAQSSKQVAAQ-----SOPQS----- 2194
Db 3720 HLGISMTLARELNFVDEIINOIIVENPNSLISQSEMLKRWYTRDKNATTDALTSVLT 3779
QY 2195 -----NWOG-----QSPVAVQSPQSRIRPST 2216
Db 3780 KINRDIIVTLLEGPIFYDIGNISGTRSPADENNWFHDPVDGPSLQVLELPTGLHTYPT 3839

```

```

QY 2217 PSOLSPGQSQVQVQTTISQPIPIQHTSLQ---IPSG-----QPQSQPOVQOSSQTOLSSQ 2269
Db 3840 PFO-----QDDYESDISIESPLTPSRSLDGLVPSQGNIEHSDGFPVYTAERASLEDK 3895
QY 2270 TLNQVSVSPSPQLOIQOPQVIAVPOLOQOVVLSQIOSQVVAQIAQOQSGVPOQIK 2329
Db 3896 LEDSVPLTE-----MPEAVNDESOLENCVLS-----YPOLIGNLAGSPKDYK 3939
QY 2330 ----LQPLQIOQSSAVQTHQIONVYTVQAASVQEOLOQVQOLRDQOQKKQOQIEKRE 2385
Db 3940 PAERKRLGVSSQEOEKSGSPDEEMBEKLSLFEQIOLEEVSESEMEEKQVQALIKR- 3998
QY 2386 HTLQASNOSEIIQOVYMKFNAAV-----IEHLQKKSMTPA--BRENORAIVCNQMYKI 2439
Db 3999 -----VQOAELEMSSTIGMONETSSGNLESCAQARVYTGILDRDSDPQCHDSTSYL 4053
QY 2440 LDKIDKEE-----KQAKKRKREESVEQ--KSKQNAATLSALLFKHKEQLRAEIL 2488
Db 4054 KGEAGKEFANGSHTEITPEAKTKSYFPESQNDVGKQSTKETLKPkihSGHVEEPASPLA 4113
QY 2489 KRALLDKQIOVEELKDKLKIKEKDLMOQAQATVAPAPCPVTPVLPAPAPPPSP 2548
Db 4114 AYQKSLSETSKLITE-----TKPCVPVSMKMSKSTSPADGK 4150
QY 2549 PPPPGVQHTGLSTPLPVAQKRRKEEKSSSKSKKMMISTSKETKKDKR 2602
Db 4151 P-----RLSLHEEGSSSGSEQKQSGCFV--KTKKEIR 4181

RESULT 7
US-09-724-676-69160
; Sequence 69160, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69160
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-724-676-69160

Query Match      2.7%; Score 388.5; DB 5; Length 3259;
Best Local Similarity 16.8%; Pred. No. 0.0096;
Matches 576; Conservative 493; Mismatches 1207; Indels 1155; Gaps 129;

QY 17 QDEDEDEDEDEDEDDSDYPREMEDEDDDASCTESSFSRSHYSTSPG-----RR 68
Db 131 KETESODDDEIEKTRDQSF-----ASIALKRRSYLTERPOMIODEYRSNK 176
QY 69 KPRVHRPSPILEKQDIPLEPKSSSEDLAVPNDHIMNVAIYEVLRNCTVLRSLPFR 128
Db 177 IPPHHLIQALILEYKTPVLD--NSSDYSAMASQVRLHPLIKFEFLYYAISFSVKINWCV 234
QY 129 EDCQAL-----VSOEQCTLMAEMHVLLKAVLREEDT-----SNT 164
Db 235 DTFSSNQHIIRRVNITFSKSIDVASPIQLITDNVPADKNGCVITISIQPGLMYPKSSS 294
QY 165 TFGPADLKDSVNSTLYFIDM-----TWPEVLRVYC--ESDKE-----YHNVLPYO-- 208
Db 295 SHGSYALKRAGIQFYVFLSNLSDSYKSVSGEVIYNYAPAPSKSNINMYSSSLPFSKIL 354
QY 209 -EADDPYGVENKIKYQLQVLDQFLTTNTAREELMSSEGYIQYDDHCYVCHKIGDL---- 263
Db 355 TSAAPLISPSLKSIVSPVKSRRVDVSISSAKITMASLSSPYKQMPGHAVALVNGSISPLK 414
QY 264 ----LCCETCSAVYHIECCVKKPPLLEVPREDEMCQCVVA-----HKVVPVTCQVAREIQ 312

```


Db 415 VASSSTLINGCATATLO-----EKISATSIVSSVSAATDVYKYSTTDCA--TTP 467
QY 313 NKPIRHEPIGYDRSRKY-----MFLN--RRLIEEDTENENKKIWYSTVOA 362
Db 468 TQVIGCFSTISFSVSKNSFRKCTLYIPWVVICNYLCNFINVNSASISVWVLPALX 527
QY 363 ELIOLCDKDYWEALCKLIEEMREEIH--RHMDITEDLTNKAAGSKSLA----- 411
Db 528 KLPDSNSFTKSAALISPLTKLTLETTERHPOHFSRTSPV-----KSLFLAASALKLSTP 582
QY 412 ---AANEETIESIRAKKGDIDNVKS-----PEETEKONENENSKOAEKRR 455
Db 583 SLSUSSQELIKVYAEKEDLMRTALOTDVPEKFPPELREKGIIDDEBPKEYEKV 642
QY 456 EEFEDQS--LEKD-----SDDK--TPDDPEQKSEVGFSEKSENGELSESP 499
Db 643 EDLVYKSEILKKDYCDVKNKSPKSPKSDKGHPEDDIEFSSV--EIREKROOAAASQSP 700
QY 500 G-----AGKAS-----GSTRITRLRNP--DSKLSQLK-----SQOV--A 531
Db 701 SUPERVQYAKAKASEKDYMLTKYVIDLTNDIGSSSLTNLKYFEDAKKKGEGGOKRVLRP 760
QY 532 AAAHEANKLFKFGKEVLYVNSQGEISRLSTK---KEVIMKGNINNYFKLGQEGKYRVYHN 588
Db 761 AIALOBHKLKMPASAKRTSTSEKELCKMADSPFGDTILES--PDQFQSOHODKSPLSOS 818
QY 589 QYSTNSFALNKHQHRDHDKRRHLAKFCITPAGEFKMNGSVHSGVLTISTRLTITOL 648
Db 819 GRETSEKTPPSAPOSACETTGPKPLPHEVPIPIVITETPTEVYVIRSPYASADVQTOP 878
QY 649 ENNT--PS-SFLH-----PNMASHRAMIKAVOMCSKPREFALALAILCAKPVVWLP 700
Db 879 EBPVSKSPPTMELEPKPTTSIKEF--KAFOM----- 911
QY 701 WREPLGHTLHMSTIEREKEKVKKKEKKEEETMOATWYKTFPVYKHQWVKOGEE 760
Db 912 -----KASSEEDHNNVLKSGMKRKEETHITTTTRVYHSP----- 947
QY 761 YEVTOGVMWISKTIVYRFPVKLPDNTVNVYRKSLSEGTKNMDEMDESDKRCRSRPK 820
Db 948 -----PGG-----EGASEREETMSVMDIMKAKQOSGR 974
QY 821 -----KIKIEP-----SEKD-----EV--KGSDA--- 838
Db 975 DESKELAGLFEHKSAVSPVHKSAETSQAQHAKEKDMQPKLERILEVHIEKQNAEPT 1034
QY 839 -----AKGADQHEMDI--SKITEKKODVKEILDSDDK--PCKEERPEVVDMMKT--ESHV 889
Db 1035 VIIRETKKHPEKEMVYQKOLSRGIDINLDFLPEKHDAFPCCSEOGQOEELTAEESLP 1094
QY 890 NCOESSQOVVNVVNSGEFHLRTSYKK-----KTKSSKL-----DGLLEHRIKQFTL 934
Db 1095 SYLESSRVN--TPVSOEEDRSPSAOLISDQSYTKLKLHOSHIEYHDELSELRGESYRF 1153
QY 935 EERKOL--EKIKL-----EGIKGICKITSTNSKNLSSESPIYTKAEGCO 977
Db 1154 AKKMLSEKILDVSHSTEESTVDHAGRPSEELQSGDRSRREKRIATAPKKEILSKIKYDVS 1213
QY 978 SPSMROESPPNNDOPPELLIGCSQSDSSVLRMSDPST-----TNKLYRK 1024
Db 1214 ENGVOG-----VSKDEHFKVTVLHVSQ-----VSSPKHAMMMRTEDRLDREGRELTLYE 1264
QY 1025 DVLVDVSIRESPTCKPKONSIENDIEEKVSDLASGOEP--TKSKTKGNDFTIDSKLAS 1083
Db 1265 DRV--DRYVKAEBEKLTEVSOEFFROKTEKLN---ELQSPREKARKKKEYSOSSQPTSS 1319
QY 1084 ADD---IGTLIKKNNK-----PLIOESDTIVSSSKALHSSVPKSTNDRD 1126
Db 1320 SEKVLLELLASNDSEWYKAROHGPDGQGFPAKEEKAPLPSPEKMYLS--OQTEDESK 1376
QY 1127 AT-----PLSRAD-----PEGLGDSSESTLENSSDVS-- 1158
Db 1377 STVEAKGSIQSQKAPDPOSGFOLKOSKLSIRLKFEGQTHAKSKDMQSDRKSQOSKRI 1436

QY 1159 ----IODS-----SEEDMIVONSNESISEQ-----FRTREODVEYLE----- 1191
Db 1437 PVKIOESKLPYQVFAREKQAKIDLPEDESYSVQKDFMWLTKTKDEHAOSNEIYVNDGS 1496
QY 1192 -----PLKEIYVSGESTGCNEDLPLVKGTEANG--KKPSQ 1224
Db 1497 DNYKORTEMSKAMPDSESOQAKDLACHITSDLTGRPMKRVFRTWSSGATNNKQ 1556
QY 1225 OKKLEERPYNKCSQJIKLKNITDKNNNE--NRESEKKGORTSTFOJNGKDNKRYLKEG 1282
Db 1557 KEKLSHVLVH-----DVRENHIGHPESKSVDOKNEMASTERRK---LLTNG 1601
QY 1283 CLKIEISERVYSGNVBPYNNINKIIPENDIKSLTVKESAIRPPIGVDYIMEDFENRNS 1342
Db 1602 SLSIEIKEMYTKSPS-----KKVLREVYVKGDPHGGILLDQPSRRSES 1644
QY 1343 ETKSH-----LLSSDAEGNTRDSLETLPSTKRESISTOTTTTPSACPSNSVNOV 1392
Db 1645 SAVSHIPRVADERRMLSSNIPDGFCEQS--APPKHELQKLSQSSMSKETVETQHNST 1702
QY 1393 EDMIEITSEYKVT-----SPITSEE-----ESNLSDNPDIDENGLP 1429
Db 1703 EDEKVTYSEISKYSKQSYVGLCPPLBETETPTPKSPDSLFEPSPKESPSDVFDHS--P 1760
QY 1430 INKNENY-----NGESKRRTVITEVTTMTSTVATESKTIYKVKGDQOTVASTENCAR 1483
Db 1761 IDLEKLAFLAQEGGEKELTLPVYV---SFVOVGQKQYEKEIQOGGVKKIISOECKTVQ 1816
QY 1484 SYVTITTTVTYKLTSTGSGVDIISYKESQKTVVTTVYDLSLTITGGILVYTSMT----- 1538
Db 1817 ETRGFTYPTTROOKQPPSPGSSPEDDTLEOVS-----FLDSSGKSPLETPETPSSEE 1866
QY 1539 VSKEYSTROKVKMLKF-----SRPKTRSGALPSYRKFPVYKSKKSLFV-- 1583
Db 1867 VSTEFTSKTPDSLIAYIPGKRSPPIPEVSESESEEOAKSTSL-----KOTIVETAYE 1919
QY 1584 --LPNDLKLKARAGIREVPYFNYPNAKPALD----- 1613
Db 1920 REMPNVDSKSNQRPANNRAYIEFPPLDADQIESDKHHYLPKEKEDMLEVNLQDE 1979
QY 1614 -----TWYRPS-----RPFITGWRRYRLOTV-----K 1636
Db 1980 HDKYQALAEPIYRQV--PSPVPGADVSDSDDESIIQVPPVKYVTKLEKVEDQEKRP 2038
QY 1637 SLAGVSLMLRLMAS-----LRMDMAKAVPPGGGSTRRETSETETTEIILKRRD 1687
Db 2039 ASAEKASNOKELESNGSGKDNFGLGIDSPQNEIAGONNDOSITPCSIATTAFFSHD 2098
QY 1688 -----VGPYGRFE-----YCIKILICPIG-----VP 1709
Db 2099 ATIEDSLDGDIDDEDGILTESDKLPLOAMEIKKDIWNTREGILKADRFSQSKLEVIE 2158
QY 1710 ETPKETPTPOKGLRNSALRPKRRETPKOTGPVILIEWAAEBELEMEIRAFERAYEK 1769
Db 2159 EEEKVPDEKPKPSKSSSEKTPDKDOKSGAOFLLLEGHNPDRSVFPPTDYFYKVIDEE 2218
QY 1770 AQAVEQQA-----KKRLQOKPVIATSTSPSTSTISPAQ 1808
Db 2219 ATEFKTVATKGLDFPWSNNRGDEVEFDSKREDETRKPGLAVEDNSP--ATTPDTTPAR 2276
QY 1809 KVVAVPISGSVTTGTGMVLTITKVSPPATVTFQONKNEHOTFATV----- 1853
Db 2277 -----TPTDESTPTSEPNPFHEHGKMEEMTRSAIDMSKADVEERLOFF 2322
QY 1854 -----KOGOSNGSVVOYQOKVLGILPSSIGTSQOTFTSFQPRATYVITPN-- 1899
Db 2323 QIGEHNTSEKSGDQGGKDSMTATPO-----POSQDTIVE--TNLERNVETPTVEPNBS 2375
QY 1900 -----TSGSGGTTNSQVITGQAIRPGMTVTRTPELOGSTLKAIIIRPVWVQOPA 1949
Db 2376 IPTSGEQEGTSSGSLKSAATNTSKYDP---KLRTPKM-----GISASITMVKKEG 2427

QY	1950	POOVMDI-----IGOPVSTAVSAPMNVSTPG-----	QKSTSVT- 1986
Db	2428	PGETTICKIEVMTSCGLENETITMTSNTANSOMGVREHEKHDFOKDNFNANNNDSDSTI	2487
QY	1987	STSNIOS-----SASOPRPPOQGVKLJYMAOLTQLTQGHG-----	GNGLJVVVIQ 2031
Db	2488	QTDIMSNIVLTEHSAPCTTEKDNFVKSNGKTKGVLOGHCVDRDKOKVLEDOOKTELLI	2547
QY	2032	GOGQ-----TGQLODLPQGVYLP	2051
Db	2548	GIRKSKLPIKATSPKDTFRPNHMTKSKKKQYQSEKTKALTITSSCDVYKSRIPKN	2607
QY	2052	GPGQOLMOA-----AMPNGTVORFETPLATTATTAATTTTTSVTA	2093
Db	2608	TPRDNITIAVRKACATQKQOGPEKAKAOLPSKLPAKVNSTCVTTTATTTATTTTTTTTT	2667
QY	2094	AG-----TGE-----QROSKLSPQOVQODTLP--	2117
Db	2668	TCTVTKRSQOLKEVCKHSIEFYKGISGETLKLVDRLSEEEKKMOSELDEEEKYLLKXV	2727
QY	2118	-----PA-----QSSVSGAPKAPQPTAOPASAPQOTQPOSPAPREVOQTOPEV	2160
Db	2728	VVRDPGWPAPFAGYKNSAROKKTEAALPLKSKSEKASEKSSKSRTOPOGPCE-----	2778
QY	2161	QTOCTTSSVHPSEARPPHAAOSSKPOVYAAQ-----SOPQS-----	2194
Db	2779	--RTDIMAIVAD--HGLSWTELARELNFSVDEINOIRVENPNSLISQSEFMLKKXV	2832
QY	2195	-----NVQG-----	Q 2199
Db	2833	TROCKNNTTALTSVYLTKINRIDIYTLLEGPIFYDGNISGTFSPADENNVRHDPVDGVS	2892
QY	2200	SPVAVQSPSOTLRPSTPSOLSFPQGOQVOTTSOPRIPIQHTSLQ-----IPSQC-----QP	2252
Db	2893	LQVLELPTPLGHTHPRPFG-----QDDYFSDISLIESPLRPSRSLDGLVSGOIEHSA	2948
QY	2253	QSOPOVQOSTLSSGOTLNVQSVSPSPRPOLQOOPQOVIAVPOLQOVLSQIQSQ	2312
Db	2949	DGPVYVTAEDASLEDSKLEDSVLTB-----MEAVDVDESQLENVCLSE-----	2993
QY	2313	VVAQIAQOAGSVPOQIK-----LOLPQIOQSSAVQHTQIONVTVQOASVQOBQIORVQOL	2368
Db	2994	-YPOYLGNLGSBPRDYKPAERPRKLYGVSDEQEKGKSGPDEEMEEKLSLEFEDIDLEGV	3052
QY	2369	RDOQOKKKQOQIEIKREHTLQASNOSEIIOQKVMKHNANV-----TEHLKOKKSMTPA--E	2422
Db	3053	ESEMEETEKOVALIKR-----VOQAELEKSSITGQONTSSGNIENSCQAARRYTGGLD	3106
QY	2423	REENORMIVQNOVWIKYLLDKIDKE-----KOAKKRKRRESVEQ--KRSKONATK	2471
Db	3107	RLDSPPOGCHDSITSYLKAGKGFANGSHTTEIPRAKTKSYFSPSODVQNGQSTKEKL	3166
QY	2472	LSALLFHKEDLRAELIIRKALLDKDLQIEVOELKRLDKIKKEDLQMLQAOATVAVAPC	2531
Db	3167	PKIHSGHVEEPPASPLAAQKSLSETSKLIEB-----TKPC	3203
QY	2532	PPVTPVLPAPAPPPSPRPVQVHTGLSTPLPVAQOKRKEEKDQSSKSKKKKMTS	2591
Db	3204	VPVSMKKMSRSPADGPR-----RLSHHEEGSGSGEQKQGEFGK	3243
QY	2592	TTSKETRKDKT 2602	
Db	3244	V---KTKKEIR 3251	
RESULT 8			
US-09-724-676-68931			
; Sequence 68931, Application US/09724676			
; GENERAL INFORMATION:			
; APPLICANT: CompuGen LTD			
; TITLE OF INVENTION: Variants of alternative splicing			
; FILE REFERENCE: 129181.4 CompuGen			
; CURRENT APPLICATION NUMBER: US/09/724, 676			

```

; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68931
; LENGTH: 3898
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68931

```

Query Match	2.7%:	Score 385;	DB 5;	Length 3898;
Best Local Similarity	17.2%:	Pred. No. 0.013;		
Matches 481;	Conservative 478;	Mismatches 1036;	Indels 808;	Gaps 114.

QY	305	DCVAEI-----QKNP- YIHHEPIGVDRSRKW-----FLNRLLIE	341
Db	113	DCSSEVNCGSFWMTGKPYMLLREBEFGVDISTEGGADSTFHEMMESLAGQHETE	172
QY	342	EDTENEKKIKWYSTKVOLAELIDL-----DKDYAEELCKLT	381
Db	173	ELNRELEBMRYATEGILOLOFEAAIKORGIITQLTANLOQARREKDETFMREFEILT	232
QY	382	EEMRE-ELH-RHMDITEDLTNKARGSKSTLANAEI-----LESTRARKG	426
Db	233	EO5OKILOFOLOASTELTNSHSTAAULLAKOOILHQOOLEBDHLLIEDYQKKE	292
QY	427	DI-----DNVSPRETEKDKNETENENDSKAEKNRE-----EPEDSOILEKDS	467
Db	293	DFTMOISLOEKIK-VYEMQDR-KVENSKEBIOEKETIIEBLNTKIIEBEKKTLEKD	350
QY	468	DDKTRP-----DDPQGSSEVGDKRSEKSNELSESQAG-KGASSTIITRLNRP	518
Db	351	KLTPADKLGLGLOLOIYOKNOETIKNNKLETJNSKOROSSBEEIOLMGTVELOKRHK	410
QY	519	DSKL-----SOLKSQOAAAANAHA-NLFLPEKREVLVUNSOGISRLSTRKEYI	566
Db	411	DSQETPDIVQMBEQDORKLEGLRAELDEMGQOYOMKELLIRQMAQMEMKTRHNKE	470
QY	567	MKGNNINFKIGQBGKTRYVHNOYSTFSLN-KHNEHDHDKRHLNAKCSLTPAEF	624
Db	471	MEMLRASYNT-----TVNEDOIKLWVAINELNKLIDTJNSQKKEKLEL-----	516
QY	625	KWNSVHSGKVLITSLTRLITLOENNIPSSFLHPMASHRAWIKAVOMCSGPREFALA	684
Db	517	---GLILEEKALOQROJLEDEVEL-----SFREQ-QIQRARQITABE-SKLWNAKS	564
QY	665	LAILECAVPVYMLPIRREPLGHTRLHRMS-----TEREKEKVKKKEKKOEBEETMOA	740
Db	565	LSTVEDLKAELIVSASESRKELELKHAEVYNYKIKLEMLEKEKNALIDMAESOEAELER	624
QY	741	TWVYTPPVKVHVKQKGE---EKRYVGYGSMWIKLTHVYRPVPLPENTVANYRKSLE	797
Db	625	LRTOLLSHEBELSKLEDELEIRLN-----TEKLNKLGIHIKOOID	668
QY	798	GTKNNMDENMP-ESDRKCSRSRPHKIKIIPDEKDEKGVGSDAKAGAD-----ONEM	847
Db	669	GLONMSQKIEFTMQFEKNDLITKONOILIEISKLOQOSLVMSKSEEMTLQINLEKEI	728
QY	848	DISKITEKK---DQVKEL-----LDSDBCKRCKEPEVY	879
Db	729	EILREBEKEKGTLOEVOELOLKTELLEKOMKEREKNDLOEKAQOALEAE-NSILKOEKKTU	787
QY	880	DDDKTESHVNCSSQOVAVNVSSEGHLTSTYKRTKTSKSLDGLLEIRIKOFTLEEOR	939
Db	788	EDMLKTIHPVSQEE-----KLITLDSIKSKSDSVAKERI-ELILEENBD	831
QY	940	LEK---IKLEGGIKIGTSTNSSKNL-----SESPVITAKEGQSGDSMRQE---Q	985
Db	832	LKOOICOLNEIEIKORNTSFAEKNFEVNYQLOEAEACLLKVDLDEBKSRRKOELEYS	891
QY	966	SPANNOQPELLIGCSOSOSSVLRMSDBPHTTNKLKLPDRVLVDVDSINSPEP-----	1038
Db	892	KIKALINE-----LHLOINPTTVKM---SSVFDEKKTVAETLEMGVY	934

QY 1039 -----KCPKNSIENDIEEKVSDLASGOEPTKSKTGNPFIDDK-LASADD 1086
Db 935 EKDTLEMERLEVTAKREKLE--LSORLSLDS--EOLKQKPEISFLMEDEKSLPEKE 988
QY 1087 ICTLICKNNKPLIOESDPTIVSSSKSALHVSVPKSTINDRAPPILSRAMPECKLGDSDS 1146
Db 989 PVSILRR-----ELEIILINHRA-----ENVOSCDTOV 1016
QY 1147 NSTLENSDPTVSIODSSEDMIVONSESISEOFTREODVEYLEPKCELVSGESTGNC 1206
Db 1017 SSLL-----DGVATMTSRGAGSVSKVNSFGEESKIMVEDKVSFE--NMVTEESKQE 1068
QY 1207 E---DRLP--VGTENANGKPPSQOKLEBRPVNKCSDQIKLAKTIDKNNENRESKQO 1261
Db 1069 QILDLHLPSVTKESSLRAQOPSENDLQ-----KELNVLSKSEONDLR-LQMEAQ 1116
QY 1262 R-----TSTFOI-----NGKDNPKFIYKGECKLSEISVVSNGNEPKNNINKII 1308
Db 1117 RCLSLVSTHYDQVREYENENEK-----KALCSLKEELIFA--QEEKIKELQK-I 1164
QY 1309 PENDIKSLTVKBSAIRPTNGD-----VIMEDFNERNSEETKSHLSSSDAEGNYRDL 1362
Db 1165 HOELQOTMKTQET-----GEGCKPLHLIGLKQKAVSEECSYFIQTLCSVLGEY---- 1213
QY 1363 ETLPSKESDSTQTTTPSASCPEBSNVNOVEDMEITSEVKVYTSPTISEESNLNDF 1422
Db 1214 -----YPPALKK-----EVNAEDKEN-SCDY 1233
QY 1423 IDENGLPINK--EAVNGESKRRKVIITEVT-----IMTSVAVRESKVIIVEKG 1469
Db 1234 ISENEPELDQARYEVOADQOENMHILNKVTEBYNKLVLQTRLSIKWGOQIDGMKLEFG 1293
QY 1470 DKQTVVSTENCAKSTVTTTTVTVTSTPGSGVDIISVKEOSKTVVTTVTDSLTIT 1529
Db 1294 E-----ENLPKE-----ETEFLSHSQ----- 1310
QY 1530 GGTLVTSMTVSEYSTRDKVKLMTKSRRPKTNSGALPSYRKRFVTKSTKSTIFLPNDL 1589
Db 1311 -----MTNLDEDIVNHSKSLSLDLEKTKLEQVOELSELISSLOOQ-----L 1354
QY 1590 KKLARKGIREVPYFNYNKAPALDIMPPSPRTFGITWRYRLQVKSLSAGVSLMLRLM 1649
Db 1355 KETEO-----NYEAE-----IHCLO--KRLQAVS----- 1376
QY 1650 ASLRMDMAKVPFGGSGSTRTESETETITTELIRKRDVGPYIGREYCIKTIICPIGV 1709
Db 1377 -----ESTVBP-----SLPVDVSVITSDAQRTWYPS-----CVKKNI--DGTI 1414
QY 1710 ETPEKETPTQKRGKLSASLRPKRPETPKOTGPYIETWYA--EEBELMEIRAFAEVEKE 1768
Db 1415 EFSGEGVKEETNIYKLEKQYOEOLEEVAVIYMSIAFAOQTELSRISGGKENTASS 1474
QY 1769 K-AQAEVOQAKRLEQOKPTVATSTSTSTSTISPAQKVMVAPIGSVYTGKMWL 1827
Db 1475 KQAHAVCOQEOHYFNEMK-----LSODQIGFOFETVDVKFEKFKPLKELGEHCKEL 1529
QY 1828 TTKVGSPTVTFQOONKFNQTEFATWYKQOSNGVVOOQVLGIPSTGTSOQFTISF 1887
Db 1530 LS-----NSDPHD-----IPESDKCVLTI-----SEKFNKSD 1556
QY 1888 QPRTATVTIRPMTSGSGGTTSSQVITGPQIRPGMTVIRPILOOSTLGAIIIRTPVWOP 1947
Db 1557 KPIFVROSITHDEISVSMDSASOLMNEOELDMROELROYOEHQAOTELLROAHMRQ 1616
QY 1948 GAPQOVMQIINGOPVSTAVSAPNVTSSPGQKSLTSATSTNIGSSASQPRPQOGQV 2007
Db 1617 EKQREQOEOQ---QEEIKRLNQLAORSIDENENLSEERVLLELEKLKOLSLAGREK 1673
QY 2008 LTMALUTLOLTOGHGNGOGLTVYIOGQGTQGO-----LQILPQGV--TVLPQGOQLMQ 2059
Db 1674 LCCELRNSSTQONGN-----ENQGEVEEQTFKREKELDRKEDVPPIILSNERALQK 1726
QY 2060 AAMPNGTVORFLFTPLATTATTAST-----TTTTVSTTAAGTGEORSKLSPOMOV 2110

Db 1727 A---NNRLLKILLEVVAKTTAAVEETIGRHVLIILDRSSKQSSASLIRSEABASVSCV 1783
QY 2111 HODKTLP-----PAOSSVCP-----AKAQOTQOPSAKPPQOPQOPQAPPE----- 2153
Db 1784 HEHTRVTDSEISPSYSGSDMPFRNDIMMSKVTEEGTELISORLYVRSFAGTEIDPEDELM 1843
QY 2154 --VQTPQEVQOTTVSSHVPSEAPTHAOSKRPQVAQS-----QPOSNVQSGPPRVQS 2206
Db 1844 LNTSRLOAAVEKLEIAISPTSSQLEHAKVTOELMRESRQKQKQATDESJKQGEELRERL 1903
QY 2207 PSOTRIRPSTPSQLSPQO-----OSQVOTTT-----SOPPIQPHTSIQIP 2247
Db 1904 HESRRAREQULAVELSKAEGVIDGADEKTLFEQOIEKTDIIDRLQELLCASNRLOELE 1963
QY 2248 SOQOPQOPQOVSSTQTLSSGQTLNQNVSPPSRPQLOIQOPQOYLVAPQLOQOYVLS 2307
Db 1964 AE-----QOOIOEBRELISR---QKAMKAEGP--VEQ-----QLQJTEKLM 2002
QY 2308 QIOSQVVAQIOAQOSGVPOQIK-LQLPITQIOSSAVQTHQIONV-----YTVQASVQEO 2362
Db 2003 KEKLEVQOQAEKVRDLOQKQVAKLEIDVEQVSRFIELEQEKTELMDLROQNALQOL 2062
QY 2363 QRYOQLRDQOQKQKQOQLEIKREHTLOASNQSEI QK-QVMAKNA-----YIEHLKQK 2416
Db 2063 EKMRKFLDEQALDREHBRDVFQOEOIKLEQOLKVVPFRFQISQHTREVEOLANHLKEKT 2122
QY 2417 ---SMTPAERENQRM-VCONVMKYILDKIDKEEQAKKRRRESVEQK----- 2463
Db 2123 DKCEILLSKQLOQDIOERNBETIEKLEFRVRELOLLEDRKHFCAVEKAPPELSLEVOL 2182
QY 2464 -----RSKQNAKTLALFLFKKQOLRAELIKKRALDK-DLQJTEVO-----EEL 2506
Db 2183 QAEKDALDKREKEITNLEQL-----EQFRELEENKMBEYQOLMQOIEQKKESTRLQEL 2238
QY 2507 KRDLKIKKEKDLMOQAATAVAACPVPVPLPAPAPPPSPPPPGVOHTGLSTPTLP 2566
Db 2239 EOENKLFKD-DMEKLGIA-----IKESDAMSTODOH 2268
QY 2567 VASQK-----KKREEKD-----SSSKSKKKMISTSKETKDKTKLJCQIKPYDSKPY 2617
Db 2269 VIFGKFAQIIQEKVEIQLDQNEQVTKLOOQKLTITNKKVIEENELIRDET-----Q 2321
QY 2618 IGCDCRQNMVHRCVGLQSEAELEIDVYQPOCOSTEDAMTVLTPLEKDYEGSLKEV--L 2675
Db 2322 IEC-----LMSDQECVK-----RNREBEIQLNEVIEIKLQOELANIGQK 2360
QY 2676 RSLQAHKMAAPLEVPDPNDAPDYGVYIKPEMDLATMEERVOYRRIYEKLTIFYADMTKIT 2735
Db 2361 TSMNAHSLS---EFADS-----LKHOLDVVAIEAKLALQOQVETANEEMTFMKNVL 2407
QY 2736 DMCRRYVNPSSPYQCAVLESFFYOKLKGFKASRSHNNKIOS 2778
Db 2408 KETNF-----KMNQLOIELFLSLKRESSEVKIOS 2436

RESULT 9
US-09-724-676-68917
; Sequence 68917, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/774, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68917
; LENGTH: 3910
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-68917

Query Match 2.7%; Score 385; DB 5; Length 3910;
 Best Local Similarity 17.2%; Pred. No. 0.013;
 Matches 481; Conservative 478; Mismatches 1036; Indels 808; Gaps 114;

QY 305 DCAVEI-----QKNP--YIRHEPIGYDRSRKKY-----FLNRLLITE 341
 DB 125 DCSEVNGCSFVMGTGPTNLRLREEFGVDYSSEGAOSP7HLEMESELAGKHIE 184
 QY 342 EDENENKKIWMYSTVOLAEIDCL-----DKDYWEAEKIL 381
 DB 185 ELNLEEMEMVYTGTEBLOQLQEFEEAIKQDGIITQLTANLOQAREKETMELFELT 244
 QY 382 EEMRE-EIH-RHMDITDILTNKAGSKNSFLAANEI-----LESIRAKG 426
 DB 245 EOSOKLOIOFQOLQASELNNSTHSSAADLLQAKOOILTHQOOLEFODHLLLEDYOKKE 304
 QY 427 DI-----DNKSPETTKDKNETENDSKDAEKNE-----FEEDSLKDS 467
 DB 305 DFTMOISFLOEKIR-VYEMQDK-KVENSKEEIOEKETIIEELNKTIIIEEKKTLLEKD 362
 QY 468 DDKTPD-----DDPEQKSEVDFKSEKSNGLSESPAG-KGASGSTRITRLNP 518
 DB 363 KLTATADKLLEBLOQIYQKNQEIKNMKTLELNSKOKEROSSEELKQLMGYEELQKRNHK 422
 QY 519 DSKL-----SOLKSOQVAAAHEA-NKLFKEGKLVVNSOGELSRSLSTKKEVI 566
 DB 423 DSOEETDIOVMEQETORKEOLELAEIDEMYGQOIVQMKOELLIRQHAQOEMKTRHKE 482
 QY 567 MKGINNYFVLGQGGKRYVHNQYSTNSPALNK--HQRHEDHOKRRHLAKFCITPAGEF 624
 DB 483 MENALRSYSNI-----TVEDQIKLMVAINELNITKLODTPNSKEKLEL 528
 QY 625 KMGSVHGSVYLITSLRLTITOLENNIPSSFLPHNASHRAMIKAVOMSKREFALA 684
 DB 529 ---GLIEEKCALQROLEDLVEEL-----SFSREQ--IQARQTIAGE-SKINEAHKS 576
 QY 685 LAILECAVAVVMLPIREFLGHTRLRHMS---IREKEKVKKKKEQOEBEETMOQA 740
 DB 577 LSTVEDIKAVIVSASESRKLELKEHEAVNYKIKLEMEKEKAVILDRMAESQEAELER 636
 QY 741 TWKYTPPVKHOWKQGE---EYRVYGYGWSWISKTHYRFPKLRGMNVAYRSLE 797
 DB 637 LRQQLFSEHEELSKLEDELEIHRIN-----TEKMDNLGHYKQOID 680
 QY 798 GTRNNNDEND--ESDKRKSRSPPKIKIPDESEKDEVKSDAKGAD-----ONEM 847
 DB 681 GLQNEHSOKIETMOFEKDNILITKONQILFISKLKLOQLSVNSKSEMTLOINELQKEI 740
 QY 848 DISKITEKK---DOYKEL-----LDSOSKPCKEEPMEV 879
 DB 741 EILROEKEKGTLEQEOVQLOLTELLEKQMEKENDLOERFAOLEAE-NSILKDEKRTL 799
 QY 880 DDDMKTESHYNCOSSQVDVAVNVEGPHLRTSYKKTKTSKLDGLLEBRRIKOFTLEKOR 939
 DB 800 EDMIKITHPIVSQEE-----KLIFLDSIKSKSKSVWEKEL-ELLIEBENED 843
 QY 940 LEK--IKLEGGIGIGICTSNSSKNL-----SESPIVTKAREGOSDSMROE---Q 985
 DB 844 LKQOCCIOLENEIEKORNTFSFAEKNEFVNYQLOEYVACLKVKDLEDLSKNQOELBYKS 903
 QY 986 SPNANNQOPEDLIQGCQSOSSVLKMSPSHNTNKLKPKRYVLDDVSTRSEP----- 1038
 DB 904 KLAALNEE-----LHLOIRINPTVYK--KSSVDEDEKTFVAELMEQEV 946
 QY 1039 -----KCPKONSINDIEEKKVSDLASGOEPTKTKTGNDFFIDSK-LASADD 1086
 DB 947 EKDTTELMKLELVTKRKLE--LSQRLSDLS---EOLKQKPGELISFLNEVKSLEPEKE 1000
 QY 1087 IGTLLCKNKKPLQOEESDITVSSSKSALHSSVSKSTNDRDATPLSRAMDEGKLGDSES 1146
 DB 1001 PVSJLRG-----ELEIILHNRA-----ENVQSCDTQV 1028
 QY 1147 NSTLENSDITVSIQDSSSEDMIVONSNEISQOFRTRQDEVEVLEPLKCELVGSGESTGNC 1206

DB 1029 SSLL-----DGVVMTSGAEGSVKNSKSFGEESKIMVEDKVSE-----NMTVGEESSKOE 1080
 QY 1207 E---DRLP--VKGEANGKKRPSQOKKLEBRPVNKCQOIKLNTDOKKNNEPSEKKG 1261
 DB 1081 QILIDHLPVYTKSESLRATOPSENDKLO-----RELNLKSEQNDLR-LQWEAQ 1128
 QY 1262 R-----TSTFOI-----NGKDNKPKIYKGECLKEISESRVSGNVPKVNINIKII 1308
 DB 1129 RICLSLYSTHVDQVREYMEKEND-----KALOSLKEBELLFA--QOEKIKELQK-I 1176
 QY 1309 PENDIKSLYKESAIRPFINGD-----VIMEDFENRNSSETSHLSSSDABGNRDSL 1362
 DB 1177 HOELIQTMKOET-----GDECKPLHLIIGLKQKAVSECSYFLOTLLSVGEY---- 1225
 QY 1363 ETLPTKESDSTOTTPPSASCPESNSVNOVEDMELEISEVKKATYSSPTISEESNSINDF 1422
 DB 1226 -----YTPALKC-----EVNAEDKEN-SGDY 1245
 QY 1423 IDENGLPINKN---ENVNGESKRTKVTIEVY-----TMTSTVATSKTYIKVEKG 1469
 DB 1246 ISENEDELQDYREYQDPQENMHTLLNKVTEEYKNKLVLQTRLSKIMGQOTGMKLEFG 1305
 QY 1470 DKQTVSSTENCAKSVYTTTTVTYKLSPTSGSDVDIISYKESKTVVTTVTYDLSLTT 1329
 DB 1306 E-----ENLPKE-----ETFEELSHSQ----- 1322
 QY 1530 GGLTVSMFVSKSEYSTRDKVYKLMKFSRPKTRSGTALPSYRKFPYTKSTKSIPLPNDL 1589
 DB 1323 -----MNLIEDVNHKSKLSLDLEKTKLEQVOELSLISSIQO-----L 1366
 QY 1590 KTLARKGIREVRYFNANARPAIDIMYPSRPRTFTITMYRLOTYKSLAGVSLMLRLW 1649
 DB 1367 KETFO-----ANEAE-----IHCLQ--KRLQAVS----- 1388
 QY 1650 ASLRMDMAAKVPPGGSGSTRETSETITTEITTEIKRQDVQPYGIRFECYCIKRIICPGVP 1709
 DB 1389 -----ESTVPP-----SLPVDVSVITESDAQRTMPPGS---CYKKNI--DGTI 1426
 QY 1710 ETPKETPTPOKGLRSSALRPKRPETKOTGPVILETVWA--EELELELRAEAREVEKE 1768
 DB 1427 EFSEFVCKEETINIVKLEQYOQOLEEYAKIVYSIAFAOQTELSTRISGKENTASS 1486
 QY 1769 K-AQAVEQAAKRLQOKPVIATSTTSPTSSTSTISPAQKVVAVIIGSVYTTGTKMWL 1827
 DB 1487 KOAHAVQOQOEHFNEK-----LSQDQIGQFQFEYVDVAFKEEFPRLSKELGHEKEL 1541
 QY 1828 TTKVSGPATVTPQONKMFHOTFATWVQOGSNGSVQVOQKVLGIIPSSGTSGQOTFTS 1887
 DB 1542 LS-----NSDPHD-----IPESKDCVLT-----SEEMFSKD 1568
 QY 1888 QPRTAVTIRPNTSGSGGTSNSQVITGPQIRGMVYIRPPLQOOSTLGKAIIRTPVAVCP 1947
 DB 1569 KTFIVKOSIHDEISVSMDASROLMLNEBQLEDMRQGLVQYQEHQOATELLQAHNRQM 1628
 QY 1948 GAOQVMTQIIRQOPVSTAVSAPNTVYSTPGQKSLNATSTSNIGSSASQPPRPOGGQV 2007
 DB 1629 ERQREDEQL---QEEIKRLNROLAQRSSIDENELVSEBERVLELELALKQSLAGREK 1685
 QY 2008 LYMAQLTQLTQGHGNGGLTVIYQOGQITGO-----LOLIQGV--TYLPPGQOLMG 2059
 DB 1686 LCEELRNSSTQTONG---ENQGEVEEQTFKEKELDRKEDVPELISNERVALOK 1738
 QY 2060 AAMPNGTVQREFLFTPLATTATYAST-----TTVYTTAAGGEGQOSKSPOMOV 2110
 DB 1739 A---NNRLKILLEVYKTTAAVEETIGRHVGLILIDSSKQSSASALIMREABASVSKCY 1795
 QY 2111 HODKTLF---PAQSSVGP-----AKAPQOTAPSRPOPTQOPSPAPQE----- 2153
 DB 1796 HEETHTRYTDESIPSYSGSDPRNDIMMWSKVTEBGELELQRLVRSRGPAGTEIDPENBELM 1855
 QY 2154 --VQTOPEVOTQTTVSSHVPSEAOPTHQOSSKPOVAQS-----QPOSNVQOGSPYRVOS 2206

Db 1856 LNISSRLAAVAKLLEAISSETSOLEHNAVOTELMRESFROKOEATESLCKOEELREBL 1915
Qy 2207 PSOTRIREPSPOLSPGQ-----OSOVQTTT-----SOPPIOPHTSLQIP 2247
Db 1916 HRESRARREGLAVELAKAEVIDYADEKTLFERQIOEKDIIIDRLBELLCASNRLOELE 1975
Qy 2248 SOGQPOPOVOSVOTLSSGOTLNOVSVSPSPOLQIQOPPOVYIAVPOLOOVVUS 2307
Db 1976 AE-----OOOIOEBEELLSR-----OKEAMKAAGP-----VEQ-----QLLOETEKIM 2014
Qy 2308 OIOSOVVAVIOAOGSGVPOQIK-LQPIQIOOSAVOTHOIONV-----VTQAAVSQOEOL 2362
Db 2015 KEKLEVOCAAEVVRDDLOQKALEIDVEOVSRFTELEQEKTELMDLRQOQALEKOL 2074
Qy 2363 QHVVOQLRDOQKQKQOQOIKREHTELQASNOSEIIOK-OVMKHNH-----VIEHLKOKK 2416
Db 2075 EKMRKFLEDBQALDREHERVPOQETQKLEQOLKVPRPPISEHQREVEQOLANHLKEKT 2134
Qy 2417 ---SMTPAERENQRT-VCONVMKTIIDKIDKEKQAAKKRRRESVQK----- 2463
Db 2135 DKCSHELLSKEDLODIOERNEIEKLEFRVRELEQALLIEDRHFQVAKPELSLEVOL 2194
Qy 2464 -----RSKONATKLSALLFKHKEOLRAETILKKRALLDK-DLOIEVO-----EEL 2506
Db 2195 QAEBRADIKKEKEITNLEQOL-----EQFRELENNKNEVQOLHMOLEIKKESSTRLOEL 2250
Qy 2507 KRDLKIKKEDLMOLAQATAVAPCPVPVLPAPRAPPSPPPRPVQHTGLSTPTLP 2566
Db 2251 EQENKLEFKD-DMEKLGLA-----IKESDMSQODH 2280
Qy 2567 VASQK-----KRREBKD-----SSSKSKKKMISTTSKTKDKTKYCIKCPYDESKFY 2617
Db 2281 VLEGFAQIIOKEVEEIDQLEBQVTKLOOQOLKITTDNKVIEEKNELIRPLET-----Q 2333
Qy 2618 ICDRCQNNYHGRGVYIIOSEAEVLIDEVYCPQOSPEDAMVLTPLTEKDYGLKRV--L 2675
Db 2334 IEC-----LMSDEQVK-----RNKEEIEOLNEVIEKLOOLANIGOK 2372
Qy 2676 RSLQAHKAMPLEVPDADYGVYIKEPMDLATMEERVORRYEKLTEFVADMTKLF 2735
Db 2373 TSMNHSLS-----EADS-----LKHQLDVYIAEKLALEQVETANENETFKKNVL 2419
Qy 2736 DMCRRYNPSDSPFYQCAEVLESFVQKLGFKFASNSHNKIDS 2778
Db 2420 KETNF-----KMNQLOTELFSLKRESSEKIOS 2448
RESULT 10
US-09-724-676-68928
; Sequence 68928, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68928
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68928
Query Match 2.7%; Score 380.5; DB 5; Length 3907;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 484; Conservative 476; Mismatches 1032; Indels 823; Gaps 116;
Qy 305 DCAVAT-----QKNKP---YIRHEPIGYRSRRKYW-----FLNRRLLTE 341
Db 113 DCSSEVNGCSFVMRTGKPTNLLRREBFQVDSYSGAQDPSPTLHLEMSSELAKQHELE 172
Qy 342 EDTENDENKKIYTYSTKYVLAELIDCL-----DKDYWEAEICLKITL 381

Db 173 ELNRELEEMRVRVYGTGLOLOQEFEAIRKORDGIITOLTANLOQARREKDETMRELELT 232
Qy 382 EEMRE-EIH-RHMDITEDJFNKARGSNKSFIAANEEI-----LESTIRAKG 426
Db 233 EOSQKLIQFPOLOASSETLRNTHSTTAADLIQAKOQILTHOQOLEEDHLLIEDYQKKE 292
Qy 427 DI-----DNVKSPEETEKDKNETENDSKDAENRE-----EFEDOSLEKDS 467
Db 293 DFTMQISFIOEKIK-VYEMEQDK-KVENSNNKEIQEKETIIEELNTKIIIEEKKTELMD 350
Qy 468 DDKTPD-----DDEQOKSEYGDFKSEKNGELSESQAG--KQASSTRITITRLNP 518
Db 351 KLTTADKLGLQELQIVQNOEIKNMKLEITNSKOKEROSSEBIKQIMQTEVELOKRNK 410
Qy 519 DSKL-----SOLSKOOVAAAHAA-NKLFEGKEVLVNSQGEISRLSTKKEYI 566
Db 411 DSQFETDIQRMEOETQKRLQRLAELDEMYGOQIYQKQELIROMHMOAMEKTRHNKE 470
Qy 567 MGNINNYFKLQOEGKYRVYHNOYSTNSFALNK--HQHREDHDKRRHLAKFCPLPAGER 624
Db 471 MENALRSYSNI-----TVNEDQIKLMVNAINELNIKLODTNSQKKEKLEEL----- 516
Qy 625 KMGSGVSGKVLITSLRLITTOLENNIPSSFLHPMASHRANWIKAVOMCSKPREFALA 684
Db 517 ---GLIEEKCALQRLQEDLVEEL-----SEFREQ--IQRARQIABOE-SKLNAAHNS 564
Qy 685 LAILECAVAPVWMLPIWREFLGHTRLHRMST-----TEREKEKVKKKKEKOEFEETMOA 740
Db 565 LSTVEDLKAEIYASASRKELELKEAEVNTYKIKIEMLEKEKNANVLDMAESQOELELR 624
Qy 741 TWVYKTFPVKHOWWKQGE--EYRVLYGQGWISKTHVYREVPRLPGNTVYNRKSL 797
Db 625 LRTQQLFHSHEELSKKEDELEIHRIN-----IEKXONLGIHYQOQID 668
Qy 798 GTRKNNDENMD--ESKRKCSNSPKIKIIEPDSKEVYGSDAKAD-----QNM 847
Db 669 GLONEMSOKEITMOPFKDMLITKQNLILEISKDLOOLVNSKSEEMTLOINELQKEI 728
Qy 848 DISKITEK-----DQVYKEL-----LDSDDKCKCEPMYV 879
Db 729 EILROPEKEKGLIEDVQLOLTELLEKQOMEKENDLOEKRAOLEAE--NSILKDEKTL 787
Qy 880 DDMKTESHVNOQESSQVQVNVVSEGFHLRTSYKKKTKSGKIDGLERRIKOFTLEKOR 939
Db 788 EDMLKIHTPVQGE-----RLIFLDSIKSKSDSVWEKEI-ELIENED 831
Qy 940 LEK--IKLEGGIGIKGICTNSTNSKNL-----SESPVITAKBCCQSDSMROF---Q 985
Db 832 LKOOCIQLENEIEKORNTFSFAEKNFENVYOELOEETACILKYKDLDESKKNQOLEYKS 891
Qy 986 SPANNDOPEDLIQCGSQSDSSVLRMSDPSHTTNKLYPDRVLDVDSIRSPT----- 1038
Db 892 KKLALNEE-----LHLQRIINPTTVKM--KSSVPEDEKTYVAETLEMBEVY 934
Qy 1039 -----KCPKONSIENTIEEKVSDLASRGOEPTKSGKNDGFIIDSK-LASADD 1086
Db 935 EKOTTELMEKLEVTYKREKLE--LSQRLSDS---EQLKQKPEISFLNEEYKSLKPEKE 988
Qy 1087 IGTLLCKNNKPLIOESDPTIVSSSKSALHSSVPKSTINDDAPPLSRAMPFEGKLCGDS 1146
Db 989 PVSILRCR-----ELEIINHNRA-----ENVOCDTQV 1016
Qy 1147 NSTLENSDVTASIODSSEDMITQONSNEISDPOFTRQEDQVEVLPLKCELSGEGSTGNC 1206
Db 1017 SELL-----DGVYTMYSRKGEGSVSKYKNSFGESKILMVDDKVSFE-----NMTVGESKOE 1068
Qy 1207 E---DLPL--VGTETANGKRPQOQKLEERPVNKCSDQITLKNTTDKNNENRESEKQO 1261
Db 1069 QILDLPLPVTWESSSLRAVQSPENDKLQ-----KELNVLKSQENDLR-LQMEQO 1116
Qy 1262 R-----TSFQI-----NGKDNKPKIYLGECLEKELSESRYVSGVNEPVVNNINKIT 1308


```

Db 1117 RICLSIVSTHVQVREYEMENED-----KALCSLKEELIFA--OEKIKELÖK-I 1164
QY 1309 PENDINGSLYKESAIRPEINGD-----VIMEDFNERNSETSHLSSSDAGCNRDLSL 1362
Db 1165 HOLELOQMÖER-----GDEGKPLHLILIGLÖKRAVSECSYFLÖTLCISLGEY---- 1213
QY 1363 ETLPSFKESDSTQTTTPSASCSPESNSVQVEDMEIETSEVKKVTSSPITSEESNLSPDE 1422
Db 1214 -----YTPALKC-----EVNAEDKEN-SGDV 1233
QY 1423 IDENGLPIPNKN---ENVNGESKRRKVTITEVT-----TMTSVANESKVTIVKEVG 1469
Db 1234 ISNEBPELODYREYQODFOENNMHTILNKVTEYKNLLVLOTRLSKIMWGQOTLGMKLEFG 1293
QY 1470 DKQTVASSTENCASKSTVTTTTTIVTKLSTPSTGSDVILISVKROSKTVVTTTTYDLSITTT 1529
Db 1294 E-----ENLPKE-----ETEFLSHSO----- 1310
QY 1530 GGTLLVTSMTVSKEYSTDKLNMKFSRKKTRSGTALPSYRKFTVKSTKKSIFLPPNDL 1589
Db 1311 -----MTNLEDDIVNKHKSLSLODEKTKLEBOYQOELESLSLQOQ-----L 1354
QY 1590 KTLARKGIREVPYFNYNAPALDIMPYSPRPFTGITTWYRLQYVKSILAGVSLMRLWL 1649
Db 1355 KETEQ-----NEYAE-----IHCLQ--KRLQAVS----- 1376
QY 1650 ASLRWDMAMAKVPPGGSGTETETSETTEITTEIIRRDVPGYGRFEGYCRKLKICPIGVP 1709
Db 1377 -----ESTVPP-----SLPDSVAVITESDAQTKMTPGS-----CYKKNL--DGTI 1414
QY 1710 ETEKETPTPQRKGLRSSALSALRPKRPETPKQTPYLIETWVA--EELEIMETRAAEVREKE 1768
Db 1415 EFSEPEVKREETINVKLLEKQYOQOELEEVAKVIVSMISAFQOTELSTRISGGENYASS 1474
QY 1769 K-AQAEQQAOKKRLQEQKPRVIATSTTSPSTSTTSTISPAQKVMVABISSVTTGTRMVL 1827
Db 1475 KOAHAVQCOQEOHYFENEMK-----LSQDQIGFQTEFVQVKEFEKFKLSKELEHGEKEL 1529
QY 1828 TTKVGSFATVTFQONKNFHQTFATWVGOGOSNGSVVOQKVLGIIPSSGTSGSOQTTSF 1887
Db 1530 LS-----NSDPHD-----IPESKOCVLT-----SEMTSKD 1556
QY 1888 QPRTATVTRPNTSGSGTTSNSQVITGPQIRPQMTVTRPILQOSTLGKAIITPVMWP 1947
Db 1557 KTFIVRSIHDEISVSMDASRLMNEQLEDMROELVROYOHOQATELLROAHMRQM 1616
QY 1948 GARPQVMTQIIRQOPVSTAVSAPNTVSTPGOKSLTATSTSNIGSASQPPRQOQOVK 2007
Db 1617 EROREDEOL--OEETIKRLNROLAORSSIDENMLVSEERVLLELEALUKOLSLAGREK 1673
QY 2008 LTMQLTQLQGHGNGGLFVIVIGOGQOTGQ-----LQILPQGV--TVLPQGQQLMQ 2059
Db 1674 LCCELRRSSSTQTONGN-----ENQSEVEEQTFKEKELDRKPDVPELISNRYALÖK 1726
QY 2060 AAMPNGTVORLETPLATATTAFTAST-----TTTTVSTAAGTGEOROSKLSPPMOY 2110
Db 1727 A-----NNRLKLILLEVVTAAVEETIGRHVILGILDRSSKSSQSSASLIMRSEAEKXSV 1783
QY 2111 HÖKTLF-----PAOSSVGP-----AKAQPTAPPSARPQOTQOPQAPPE----- 2153
Db 1784 HEETHTVDESIPSYSSGSDMRINDNMMSKVTEGETELSORLVRSRAGTEIDPENELM 1843
QY 2154 --VOTOPEVQOTVSSHVSEAOPTHAOSSKPOVAOS-----QPSNVQOGSPVAVOS 2206
Db 1844 LNISSRIQAAVEKLELAISETSOLEHAKVQTEPLMRESRÖKOEALIESLQOEHLPERL 1903
QY 2207 PSQTRIPSTPSQLSPOQ-----QSOVQTTT-----SQPIPIQPTSLQIP 2247
Db 1904 HEESRAREQJLAVELSKAEGVIDYADEKTLFERQIQEKTIDIDLQELLCASNRLOELE 1963
QY 2248 SÖGQPOQPOVQOSTQTLSSQOTLNQVSVSPSRQOLQIQOPQVQIYAVQLOQOQOVVLS 2307
Db 1964 AE-----QOQIQEERELLSTR-----ÖKEAMKAEGAP--VBO-----QLQÖETEKLM 2002

```

```

QY 2308 QLOSOVVAQIOAQSQGVPOQIK-LQPLQIQOSSAVQTHQIONV-----TVQOASVQEOQL 2362
Db 2003 KEKLEVOCAQAEKVRDLOQAVALEIDVEQVSRFTELEQENKTELMDLRQONQALERQL 2062
QY 2363 QRVQQLRDQOQKQOQOIEKREHTLQASNOSEITQK-QVAKHNA-----VIEHLQÖK 2416
Db 2063 EKMRKFLDEQALDREHERHVRFOOEIQLEQOLKVVRFPQISPISHQTRVEQOLANHLKEKT 2122
QY 2417 --SMTPARRENOQRMVLQVNOVKYTLDKID--KEKQ-----AKKREESV 2460
Db 2123 DKCSLELLSKEQLÖRDI--QERNEIEKLEFRVRLQALVLSADTFQKVEDRKHFQAV 2179
QY 2461 EOK-----RSKONATKLSALLFKHKEQLRAEILKRRALLDK-DLQIEV 2502
Db 2180 EAKPELSLEVOQLAERDAIDRKKEKELTNLEQOL--EQFRELEKKNQEVQOLHMQLEI 2235
QY 2503 Q-----EELKRIKIKKEDLMQLOAQYAVAPCPVTPVLPAPRAPPSPPPPPGV 2554
Db 2236 ÖKKESTTRLOELBOENKLEFD-DMEKILGLA----- 1 2265
QY 2555 QHTGLSTPTLPVYASÖK-----RKREEDK--SSSKKKKKMISTSKETKDTKLYC 2605
Db 2266 KESDAMSTQDQHVLFGRFAQIIOEKEVEIDQLENOVYTKLOQOLKITTDNKKVIEKNELIR 2325
QY 2606 ICTPYDESKFYIGRCQNMVHGRVCVGLQSEAEILIDEVYQOCOSTEDAMTVLPTE 2665
Db 2326 DLET-----QIEC-----LMSQÖECVK-----RNREIEIQLENEVIE 2357
QY 2666 KDYEGLKRV--LRSLQAHKMAFLEVPDNDAPDYGVYIKEMPDLATMEERYÖKRYRK 2723
Db 2358 KLOQELNALIGOKTSMNAHSLS-----EADS-----LKQDLVDVIAEKLALQEQVET 2404
QY 2724 LTFEVDMTKIFDNCRYNPSDSPFYQCAEVLESFVQKLGKFKASRSNNKLOS 2778
Db 2405 ANEEMTFMKVNLKETFN-----KKNQLOLÖELFSLKRERESVEKIOS 2445

```

RESULT 11
US-09-724-676-68916
; Sequence 68916, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68916
; LENGTH: 3919
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68916

Query Match 2.7%; Score 380.5; DB 5; Length 3919;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 484; Conservative 476; Mismatches 1032; Indels 823; Gaps 116;

```

QY 305 DCVAET-----QKNRP--YIRHEPIGYRSRRKW-----FLNRLLITE 341
Db 125 DCSSEVNGCSFVNRGTGKPTMLREEEFGVDYSQEGADSPTHLMESELAGKÖHEIR 184
QY 342 EDRENEKEKKIYVSTRVOLAEILDC-----DKDYWEALCKIL 381
Db 185 ELNRELEMRVYTGTEQLOQÖFEPAIKORQDGIITQLANLOQARREDETMRRELELT 244
QY 382 EEMRE-EIR-RHMDITEDILTNKARGSKSEFLAANEI-----LESIRAKKG 426
Db 245 EQSQKQIQIQFOQLOASPTLNSTHSSYAADLLQAKQOILTHQOOLBEQDHLLEDYÖKKE 304
QY 427 DT-----DNVKSPEETKDKNENENDSKQAEKNRE-----EFEQSLKXS 467

```


Db 305 DFTMOISFIOERIK-VYEMEOOK-KVENSNIKEIOEKETIIIEELNKKIIEEKKITIELKD 362
Qy 468 DKTFD-----DDEOQSEVGDPKSEKSNELSESPGAG--KGASGSTRITIRLRNP 518
Db 363 KLTJADKLIGELOEQIVQKNQKIKNMKLELTNSKQKROSESEETKOLMGIIVELQORNRKH 422
Qy 519 DSL-----SOLKSQVAAAHBA-NKLEKEKEVLVNSQGEISRLSTKREVI 566
Db 423 DSOFTEDIQORMEQFORLEQIRAEIDEMVGGQIVQMKOELIRQMAOMEKKTIRHKG 482
Qy 567 MKNINNYFKLOGEGKYRYVHNOYSTNSPALNK--HOHREDHKKRHHLHAKFOLTFAGER 624
Db 483 MEALMSYSNI-----TYNEDOIKLMAVAINELNKLDDTNSQKKELEEL----- 528
Qy 625 KMGVSHGSKVLLITSLRLITOLENNIPSSFLHPMASHRAMVAYOMCSKPREFALA 684
Db 529 ---GLILEEKCALQORLEDLVEEL-----SPFSREQ--IQRAQOTIABEO-SKLNBAHKS 576
Qy 685 LALIECAVAPVWMLPIWREFLGHTLRHMTS-----IEREKEKVKKKEKKOEPEETMOQA 740
Db 577 LSTVEDLKAIEIVASBSRELELKHAEVYNYKIKLEMLEKEKNAAVLDRABEQAELER 636
Qy 741 TWVKYFEPKHOVWKOGE---EYRYTGYGMSWISKTHVYRFPRLPGNTVNYKRSLE 797
Db 637 LKQOLLFSHEBELSKLEDELEIHRIN-----IEKLDNGLGIHYKOQID 680
Qy 798 GTRNNMDND--ESDKRCSRSPPKIKIEPDEKDEKVGSDAAKAGD-----QNEK 847
Db 681 GLONEMSQKIEFMQFEDKDNLTRKQNLJIEISKLDQOSLVNSKSEMIQLQINELQKEL 740
Qy 848 DISKITEK-----DQVYEL-----LDSQDKPCKEEMVEY 879
Db 741 ELTROBEKEKGLQEOVQLOKTELLEKOMKEKENDLOEKAQLEAE-NSILKDEKCTL 799
Qy 880 DDMKSESHVNOCESSQOVVWVNSEGFHLRTSYKKKTKSKLDGLBERRIKOFTLEKOR 939
Db 800 EDMKLTHTVSOE-----RLIFLDSIKSKSDSVWEKEL-ELLIEEND 843
Qy 940 LER--TKLEGGIKGIGKISTNSKNL-----SESPVITRAKEGQSDSMROE---Q 985
Db 844 LKOQCIQLEBIEKQNTSPFAEKNFEVWQELQOEYACLLAKDLEDSKKNQOBELEYKS 993
Qy 986 SPANNDOPEDLIQCGSQSDSVLRMSDPSHTTNKLYPKDRVLDDVSIRSPET----- 1038
Db 904 KKLALNEE-----LHLQIRINPTIVKM--KSSVDEDEKTFVAETLEMEGVV 946
Qy 1039 -----KOPKONSIENDIEEKVSDLASGOEPTSKIRKGNDFIDDSK-LASADD 1086
Db 947 EKDTLEMEKLEVTYKREKLE--LSORLSDLS---EOLKQPEGISFLNEEVSILKPEKE 1000
Qy 1087 IGTLLCKNNKPLIOESDPIVSSSKSALHSSVPKSTNDBATPLSRAMFEGKLGCDSES 1146
Db 1001 PVSILR-----ELLEIINHNR-----ENVOCDTIOV 1028
Qy 1147 NSTLENSDPTVSIQDSSEDMIVQNSNESISQOFTREODVEVLEPLKCELVGSESTGNC 1206
Db 1029 SSLL-----DGVYTMTRSAGESVSKNKSFGESKIMVEDKVSFE---NMVTEGESQOE 1080
Qy 1207 E---DRLP--VKGTEANGKPKSQOKKLEERPYNKQSDQIKLANTYDKKNNENRESKQO 1261
Db 1081 QILDLHLPSVTRESSLRATQPEENDLQ-----KELNVLKSEQONDR-LQMEAO 1128
Qy 1262 R-----TSTFOI-----NGKDNKPKIYLKGECLKEISERSVASGVNEPYNINNKII 1308
Db 1129 RUCLSIVSTHVQYREYENKED-----KALCSLKEELIRA--OEKIKELQOK-I 1176
Qy 1309 PENDIKSLTVKESAIRPFTNGD-----VIMEDFNERNSSETRKSHLSSSDAEGNROSL 1362
Db 1177 HQLEIQTMKTQET-----GDEGKPLHLLIGLKQKAVSEECSYFLQITCSVLGER----- 1225
Qy 1363 ELLPSTKESDSTQTTTPASCSPESNSVNOVEMELETSYVKKVTSPTISEESNLSNDF 1422
Db 1226 -----YTPALKC-----EYNAEDKEN-SCDY 1245

Qy 1423 IDENGLPIRINK---ENVNGESKRTVITEVT-----TMTSTVAFESKTVIKERK 1469
Db 1246 ISENEDEPQDQYEVAYODPOENNMHTLLNKYTEEYKNLVLQOTRLSKIMWQOQTGMKLERG 1305
Qy 1470 DKOTVVSSTENCAKSTVTTTTTIVTKLSTPSTGASVDIISVKQSKTVVTTVTDLSLTTT 1529
Db 1306 E-----ENLPEK-----EPEFLSIHQ----- 1322
Qy 1530 GGLVTSMTVSKREYTRDKVTKLMKFSRPKKTRGCTALPYSRRKVTSTKRSIFVLBNDL 1589
Db 1323 -----MTNLEIDIVNHKSLSSLODEKTKLEBOVOELESLSLQOO-----L 1366
Qy 1590 KKLARKGIREVYFNYNAKPAIDIMYPSPREFTGIMTKRYRLQYKSLAGVSLMLRLM 1649
Db 1367 KEPEQ-----NYEAE-----IHQO--KRLQANS----- 1388
Qy 1650 ASLRMDMAKAPPGGSGSTRETSETETITTEIIRKRDVPGIGIREYCIRKICPIGV 1709
Db 1389 -----ESTVPP-----SLPVDVSVITTESDAQRTWPGS-----CVKKNI--DGTI 1426
Qy 1710 EPEKETPQORGLRBSALRPRKREPRTKQTPVITETWYA-BEELMLMIRAFABEVEKE 1768
Db 1427 ERSGEFVKEETNTVLEKQIOEQLBEEYAKVISMJAFAQOTELSRISGKENTASS 1486
Qy 1769 K-AQAVEQOAKKRLQOQKPTVIATSTTSPTSTSTISPAOKVWVAPISGSVTTGTRMYL 1827
Db 1487 KQAHAVCQOQOHFNEMK-----LSQDQIGQFQFETVEDVKFKFEFRLKELGHEKELL 1541
Qy 1828 TTKVSGPATVTFQOQKNFHOTFATWVKQOQNSGVVOQKVLGILIPSTGTSQOFTSF 1887
Db 1542 LS-----NSDPHD-----IPESKDCVLT-----SEEMFSKD 1568
Qy 1888 QPRTATVTRPNTSGSGGTTNSQVITGPQIRGMVITRPLQOOSTLKGALIRTPMYNP 1947
Db 1569 KTFIVQOSIHDEIIVSSMDASROLMEQOLEDMROELVRQIOEHQOATELLQAHMRQM 1628
Qy 1948 GAPQOVMTOIIRQPVSTAVSAPNTVSTPGOKSLTASVSTNSIQSSASQPPRPOGOVK 2007
Db 1629 ERQREQOQL---QEIRIKLNRQLAQRSSIDNENLVSENERVILEBELMLKOLSLAGREK 1685
Qy 2008 LTMALQTLQGHGNOGLTVVIOGQGTGQ-----LQILPGV--TVLPGGOQLMQ 2059
Db 1686 LCELRNNSSTQONGN-----ENQGEVEEQTFKEKLEDRKEDVPPELISNERVALQ 1738
Qy 2060 AAMPNGTVQREFTPLATTATTAFT-----TTTTYSTTAAGTGEOROSLSTQOMQ 2110
Db 1739 A--NNRLKILLEVYKTTAAVEETIGRHVLTIDLRSSKSQSSASLIMRSEABASVSCV 1795
Qy 2111 HODKTLP-----PAOSSVGP-----AKAOPQTAQPSARPOQTPQPSAQPE----- 2153
Db 1796 HEHTHVTTDESIPSTYSGSDMFRNDIMMSKVTEEGELSLQRLVRSFAGTEIDPEHEELM 1855
Qy 2154 --VOTQPEVOTQTVVSSHVPSAOPTHAOSKROVAOS-----QPSNVQOSGPVYOS 2206
Db 1856 LNISSRLQAAVERKLEBAISETSSQLEHAKVQTELMRESFROKQEATFESLKQOELERL 1915
Qy 2207 PSQTRIRPSTPQSLSPQ-----OSOVOTT-----SQPIPIOPHISLQIP 2247
Db 1916 HESRARQOLAVELSKABEVIDGAYADEKTLFERQIOEKETDIIIDRLQOELLICASNRLQELE 1975
Qy 2248 SOGOPOPOPOVOSTQTLSSGOTLINOVSVPSPRPLOLOQOPQVIAVAPOLQOQOYVS 2307
Db 1976 AE-----QOQIOBERELLR---QKAAKAEKGP--VEQ-----QLQOETEKLM 2014
Qy 2308 QIOSQVVAOIQAOQSGVPOIK-LQPLQIOOSAVOTHOIONV---VTVQAAVQEOQL 2362
Db 2015 KEKLEVOQCAKEVVRDLOQVQKALEIDVEQVSRFTELEQKNTELMDLRQONQALEKL 2074
Qy 2363 QRYOOLRDQOQKKQOQOIEIKREHTLOASNOSEIIOK-OVMKHNH-----VIEHLKOK 2416
Db 2075 EKMRKFLDEQALDREHERVDFOEOIOLKLEOOLKVVPRFQISEHOTREVEOLANHLKERT 2134

QY 1283 CLKESRVSNGVNEPKVNNIKIIPENDIKSLTWKESAIRPFINDVIMEDFENRNS 1342
 Db 1593 SLSEIKEMTKSPS-----KVLVREYVYKGDHGGGLDDPPSRSES 1635
 QY 1343 ETKSH-----LLSSDADAGNRSLSLETLPSTKESDSQTQTPPSACSPESNSVNOV 1392
 Db 1636 SAVSHIPVRADERRMILSSNIPGFCES--AFPKHELKSLQSSSKKTETVETQHNSI 1693
 QY 1393 EDMEIETSEVKVY-----SPTISEE-----ESNLNDIDENGLP 1429
 Db 1694 EDEKVYVSEISKVSKHQSYGLCPLEETETSPTKSPDSLSEFSGPKESPSPDVFDHS--P 1751
 QY 1430 INNENV-----NGESKRTVITEVTMTSTYATSEKTYIKVKGKQIVVSTENCAR 1483
 Db 1752 IDGLEKLAPLAOTGEGKEITLPPVY--SEVOVGKQYKEITQOGGVKKIISOECTVQ 1807
 QY 1484 SYTTTTTTLSTPSTGSDVILSVKESQKTVVTTVTDLSLTGGTLVTSMT----- 1538
 Db 1808 ETNGTFTYTTQOQOPSPQSPQSPEDDTLEQVS-----FLDSSCKSLTTPETPESSE 1657
 QY 1539 VSEKYSTRODKVKLMKT-----SRPKTRSGTALPSYRKFTVTKSTKRSIFV- 1583
 Db 1858 VSEYFTSKPDSLIAVYIPKPSPIPEVSESESEEQAKSTSL-----KQTTVEETAVE 1910
 QY 1584 --LPNDLKKLAKKGGIREVPYFNNAKRALD-----1613
 Db 1911 REMPNVSKDSNQRPNNNRVAYLEPPPLADQJIESDKHHYLPKEVDMIEVNLQDE 1970
 QY 1614 -----IMPYSP-----RPTGIMRYRLQTV-----K 1636
 Db 1971 HDKYLAEPIYIRQP--PSPVPPGADVSDSDSDSIQPVVVKYTKKLKLVDDQKKPK 2029
 QY 1637 SLAVSILMLRLMAS-----LRMDMAKAVPPGSGSTRETSETTEITTEIIRRD 1687
 Db 2030 ASAEKASNOKELESNGSKDNFGLDSPQNLQNGNNDQSTESLATTAEFSDTD 2089
 QY 1688 -----VGPGLIFE-----YCRKLTICPIG-----VP 1709
 Db 2090 ATEIDSLDGLDDEDDGLTESDSKLPQAMEIKDWMTEGILKPADRSFSOSKLEVIE 2149
 QY 1710 ETPKETPTQORGLRSSALRPKRETPKOTGPVLIETWVAEELLEMEIRAFERVEKEK 1769
 Db 2150 EEEKVAPDEKPPKSSSSSKETPKDKDOKSGAOFPLEGHNDRSVPTDYFYKYADEEF 2209
 QY 1770 AQAVEQA-----KKRLQOKPTVIATSTSPUSSTSTISPAQ 1808
 Db 2210 ATEPKTVAATGDFDPMNNRGDEVEFDSKRDETKPGLAVEDSP--ATPDTTPAR 2267
 QY 1809 KVVVAATISSVTTGTAMVLTTKVGSALYVTPQONKNPHOTPAWV----- 1853
 Db 2268 -----PTDESTPTSEPNPPFPHGKMEEMTRSGAIDMSKRDVFEERLOFP 2313
 QY 1854 -----KOGSNSGVVVOQKVLGIIPSSGTSGOQTSTQOPRAVYTIRPN-- 1899
 Db 2314 QICEHTSEKSGDQGGGDSMTVATPQ-----PQSGDTTVE--TNLERVETPVEPNPS 2366
 QY 1900 -----TSGSGGTTNSQVITGPQIRPGMTVIRTPLOOSTLTKALIRTPVMOPGA 1949
 Db 2367 IPTSGCOGSETSSGSLKSAATNTSKVDP--KLRTPIKM-----GISASMTMKKEG 2418
 QY 1950 POOVMTQT-----INGQPVSTAVSAPNTVSSTPG-----OKSLTSAT- 1986
 Db 2419 PGKITIKIEAVMTSCQGLENETITMTISNTANSQGVPRPEKHDFOKDNFNNNNNLSSTI 2478
 QY 1987 STSNIOS-----SASQPPRQOGOVKLMQMLQTLQJGHS-----GNGGLLVAVJO 2031
 Db 2479 QTDNIMSNIVLIEHSAPTCTTEKDNPKVSSGKTGVLQGHCVRDQKVLGEQOKTKELI 2538
 QY 2032 GOGQ-----TTGQOLIPQGVTLV 2051
 Db 2539 GTRQSKLPIKATSPKDTFPNNHMSNTKASKMKQVOSSEKTKALTLTSSCVDVAKSRILPVKN 2598
 QY 2052 GPGQOLMQA-----AMPNGTVORFLFTPLATPATATASTTTTIVSTPA 2093

Db 2599 TPRDNITIAVRKACATQKQGPPEKGAQOLPSKLPVKVRSCTVTTTTTATTTTTTTTTT 2658
 QY 2094 AG-----TGE-----QROSKLSPMQVHODKTLV--- 2117
 Db 2659 TSCIVAKRSQLEKVCNHSLETFKGISGELTKLVDRLSSEKKMQSLSPBEKKYLLKHHV 2718
 QY 2118 -----PA-----OSSVGPAAKQOPQTAOPSARPQOTQPSAPQEVQTOPEV 2160
 Db 2719 VWRDFPGWPAFGYKKSARDKTEAPLKSKEKAGSRKSSRRTPGQSPCE----- 2769
 QY 2161 QTOYSSHVPSAOPTHAASSKPOVAAQ-----SQPOS----- 2194
 Db 2770 --RTDIRMATVAD--HGLSMTELARELNFVSVEIDNOIRVENPNSLISQSEMLKKVY 2823
 QY 2195 -----PA-----NVOS-----Q 2199
 Db 2824 TROGKNATTPALSVLTIKIRIDIVTLLEBPIDYGIISTRTSFADENNVFHPVDGYPS 2883
 QY 2200 SPVRVQSPQTRIRPSTPSQLSPQOSQVQTTSQPIPIQPHNTSLQ--IPSGC---QP 2252
 Db 2884 LOVELETPTGLHTPPPPFQ---QDDYFSDISSIESPLRTPSRLSDGLVPSQGNIEHSA 2939
 QY 2253 QSQPOVQSSQTOLSSGOTLNOVSVSSPRQOLOIQOPQOVIAVPQLOQOVYLSQIQSQ 2312
 Db 2940 DGPPVYTAEDASLEDSKLEDSVPLTE-----MPRAVVDSEQLLENVCLSE--- 2384
 QY 2313 VVAOIAQOQSVPOOLK---LOLPQIOQSSAVOTHQIQNVTVQVAAVQEOLOVRQOL 2368
 Db 2985 -YPOYLONLGSFRDYKPAEPKRLGVSSQOEKKGSPDEEMEEKKSLFEDIOLEBGV 3043
 QY 2369 RDOQOKKKQOQIIEIKREHTLOASNOSEILQOVVMKHNVA---IEHLKOKSMTPA--E 2422
 Db 3044 ESEEMTEEEKVQALIKR-----VQQAELMSSTIGMONEISSGNLSCAARVYTGGLD 3097
 QY 2423 REENORMIYONVMKYILDKIDKE-----KQAKKRKEESVEQ--KRSKONATK 2471
 Db 3098 RLDDSPDOCDSTSYLKGAGKFEANGSHTEITPEAKTKSPYSPESONVDGKOSTRETLK 3157
 QY 2472 LSLALFKHKEQLAEILKRAKLLDKDQIEVQBELRKDLIRKEKMLQMLQATVAAAPC 2531
 Db 3158 PKIHSGHVEEPASPLAAVOKSLETSKLITE-----TKPC 3194
 QY 2532 PVTYPLVAPAPAPSPPPPGVQHTGLSTPILPVAQOKRKRKEEKSSSKKKRKMIS 2591
 Db 3195 VPSVMMKMSRTSPADKP-----RLSHLEEGSSSSBQKQOGEGR 3234
 QY 2592 TTSKETKKDTK 2602
 Db 3235 V---KTKKEIR 3242

RESULT 13

US-09-724-676-68911
 ; Sequence 68911, Application US/09724676
 ; GENERAL INFORMATION:
 ; APPLICANT: Compugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Compugen
 ; CURRENT APPLICATION NUMBER: US/09/724,676
 ; NUMBER OF SEQ ID NOS: 97222
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 68911
 ; LENGTH: 3852
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676-68911

Query Match 2.6%; Score 379; DB 5; Length 3852;
 Best Local Similarity 17.2%; Pred. No. 0.016;
 Matches 481; Conservative 477; Mismatches 993; Indels 850; Gaps 118;

QY 305 DCVAEI-----QKNP--YIHEPIGIDRSRRKYW-----FLNRLLIE 341
 Db 113 DCSEVVGCFVNRKTCFPTMLRBEERGVDDSVSEGAQSPHLEMESELGAKOHEIE 172
 QY 342 EDENENEKIMWYSTKVLAELIDL-----DKDYWEALCKIL 381
 Db 173 ELNRLEEMEVVTGTEGLOLOEFAAIKORDGIITOLTANLOQAREKDETMEEFELT 232
 QY 382 EEMRE-EIH-RHMDITFDLNRKARGSKNSFLAANEI-----LESIRAKK 426
 Db 233 ESOCKLOLOFOQLOASTLKNSTHSSAADLOAKQOOLTHQOOLEBODHLEBYOKKE 292
 QY 427 DI-----DNVKSPEETEKDNETENDSKDAEKRE-----EFEDOSLEKDS 467
 Db 293 DFMQISFLOEKIK-YVEMQDK-KYENSNMKEIOEKETIIEELNTIIEBEKKTLELD 350
 QY 468 DDKTPD-----DDPEQGSVEYDFKSEKNGELSESPAG--KGASGSTRITTLRNP 518
 Db 351 KLTPADKLGLLOEJOIVOKQOIKNMKLELTNSKOKEROSSEIKQIMGVVEELOKRNHK 410
 QY 519 DSKL-----SOLKSOYAAAHHA-NKLFKEGKENVLYNSOGELSRLESTKREVI 566
 Db 411 DSOEPTIYOMEOETORKELOLRAELDEMYGOOIVOMQOLIKQHAQHEMKTTRKGE 470
 QY 567 MGNINNYFKLOGEGKYRVYHNOYSTNSFALNK--HOHRDHDKRRLHAKFCLTPAGEF 624
 Db 471 MENLRYSNI-----TVNEDOIKLMMNVAIINELNIKLODTNSQEKLEEL----- 516
 QY 625 KMWNSVSGKVLITSTLRLLITTOLENNIPSSFLHPNASHRAVIKAVQMSKPREALA 684
 Db 517 ---GLILEECALOROLEDELYEEL-----SFSHQ--IGORAROTIAEQ-SKLENAHS 564
 QY 685 LAILECAVKPVVMLPIWREFLGHTRLHRMTS-----IEREKEKKKKKEKOEETMOQA 740
 Db 565 LSTYEDAKAELIVSASESRKELELKEHAEVYNYKIKLEMLERKNAVLDRAESQELER 624
 QY 741 TWAKYTFPVKHQWKQGE--EYRVTYGSGWMSIKTHVRYEVPKLPNGTNNVYRSLE 797
 Db 625 LRTQOLFSEHEELSKLEDEIEHRIN-----LEKLDNIGIHYKOID 668
 QY 798 GTKNMNMND--ESDKRKSRSRPKIKIEPDSKEDVKSDAKAGD-----ONEM 847
 Db 669 GLOEMSKQKLEIWOFEKDNLTIKOMOLILELSKLDLOQSLVNSKSEMTLOJINLOKEI 728
 QY 848 DISKITEKK--DQDYKEL-----LDSDSKPCKEEPMYV 879
 Db 729 EILROEKEKGTLEQEOVLOKTELLEKQKKEKENDLOEKFALQLEH-NSILKDEKTL 787
 QY 880 DDDKTESHAVCOESSOVYVNVSEGFHLRTSYKKKTSSKLDGLERRIKOPTLEEKOR 939
 Db 788 EDMKIKHTPVSOE-----RLIFLDSIKSKSDSVWEKEI-EILIEENED 831
 QY 940 LEK--IKLEGIGICKTSTNSSKNL-----SESPVITTKAKEGCOSMSRQE--Q 985
 Db 832 LKQOQLOLENEIEKORNTFSEFAEKNEFVNYOEOLEEYACILKYADLLEDSKNOQLEYS 891
 QY 986 SPANNDQPEDDLQGCQSQSSSVLRMSDPHTHTNKLTKPRVLDVDSIRSEP----- 1038
 Db 892 KRALANE-----LHQRINPTIYKM--KSSVPEDKTFPAETILEMEGV 934
 QY 1039 -----KCPKONSJENDIEEKVSDIASRGQEPFKTKTGNDFIDSK-LASAD 1086
 Db 935 EKDTTEMEKLEVTKREKLE--LSQRLSDLS--EOLKQKPGRISFLINEEVSILKPEKE 988
 QY 1087 IGTILCKNKKPLIOESDITVSSSKSALHSSVPRKSTNDROTATPLSRAMDPEKIGCSES 1146
 Db 989 PVSILRCR-----ELEIILHNRA-----ENVQSCDQYV 1016
 QY 1147 NSTLENSDSTVSIODSEEDIMIVONSJESIEOPRTREDOVEYLEPLKCEVSGESGNC 1206
 Db 1017 SSLL-----DGVYMTSGAGAGSVSKVKSFGESKIMVEKVSFE--NMVYGEESKOE 1068
 QY 1207 E---DRLLP--VKGTANGKRPSQOKLEERPVNKCSDOIKLKNTTDKKNNEENSEKKQ 1261

Db 1069 QULIDHLPSVTKESSLRATQPSENDKLO-----KELANVLEKQENDLR-LQWEAQ 1116
 QY 1262 R-----ISTQOI-----NGKDNKPKIYKGBELKEISESRVSGNVGPKYANNINKIT 1308
 Db 1117 RICLSLYSTHVOVOREYEMENED-----KALCSLEELIFA--QOEKIKELQK-I 1164
 QY 1309 PENDIKSLVKEASAIRPFIND-----VIMEDENENSSETKSHLLSDAGNVDLSL 1362
 Db 1165 HOELQTMKTQET-----GDEGKPLHLILGLQKAVSEGCYFLOTLSVIGEY----- 1213
 QY 1363 ELPSTKESDSTQTTTPSACPSNSVNOVEDMETSEVKVYTSPTJSEESNLSNDF 1422
 Db 1214 -----YFPALC-----EVNAEDKEN--SGDY 1233
 QY 1423 IDENGLPINKN-----ENVNGSKRTVITTEY-----YMTSVATSKVIYVYKRG 1469
 Db 1234 ISENEDELQDYREYVODFENMHTLLNKYTEEYKNLLVQTRLKTIWGOQDGMKLEFG 1293
 QY 1470 DKQTVASTENCAKSTVTTTTVTYTKLSTPSTGSDIISVKEQSKTVVTTVTDLSLTT 1529
 Db 1294 E-----ENLPKE-----EFELSIHQ----- 1310
 QY 1530 GGLVLTSMVTSKEYSTRDKYKLMKFSRPKTRSGTALPYTRKTVYSTKKSIFYLPNDL 1589
 Db 1311 -----MTNLEIDIVNKKSLSIQDLEKTKLEQVOELESIISLQOO-----L 1354
 QY 1590 KKLARKGIGREVPFNFNARPAIDWYPPRPRTFGITMRYRLQOTVYSLAGVSLMLRLM 1649
 Db 1355 KETBO-----NTEAE-----IHCLO--KRLQAVS----- 1376
 QY 1650 ASLRMDMAKVPBGSGSTSTETSETITTTTELIRRDVPGYGRIFBYCJRTKICPGRV 1709
 Db 1377 -----ESTVPP-----SLPVDVVLTESDAQTMRPGS-----CVKKNI--DGTI 1414
 QY 1710 ETRKETPTPOKGLRSSALPKRPETPKQGPVYIETWVA--EELLEMETRAAERYEKE 1768
 Db 1415 EFGSEFVKKETINIVKLLLEQYOQOLEEYAKVIVSINSIFAQOTELSRISGKEENTASS 1474
 QY 1769 K-AQAVQOAKRRLQOQKPRIVATSTSPSTSTISPAQKVAVAPISGVTGTGKVL 1827
 Db 1475 KOHAVQOQOEHFENEK-----LSQDQIOGFTEYDVYFKKEFKLSLELGEHKEIL 1529
 QY 1828 TTKVSPATVTFQONKNFOTFATVWKQSGNSGVQOVQVLGIIIPSSGTSGOQFTFS 1887
 Db 1530 LS-----NSDPHD-----IPESKCOVLTI-----SEBMSKD 1556
 QY 1888 QPRTAVYTIRPNTSGSGGTTNSQVITGPQIRPQMTVIRPILOOSTLGKAIIRTPVWQV 1947
 Db 1557 KTFIVROSIDELISVSMDSRQIMLEBQLE-----DMROE-----LVRO 1597
 QY 1948 GAPQVMTQIIRGQPVAVSAPRTVASTPGQSLTSATSTNSIQSSASQPPRQOQVYK 2007
 Db 1598 YOEHQATELDR-----QAHMROMEROREQEO 1625
 QY 2008 L--TMAOLT-OLTOGHG--NOGLITVYIOGOGTTGOLIPQV--TVLPBGQOLMQAA 2061
 Db 1626 LOEETIKRLNQAQONENMG--EVEEQTFRKEKLEDRPEDVPRILSNERALOKA- 1681
 QY 2062 MPNGTVORFLFTPLATTATTA--TTTIVSTTAAGTGEDROSKLSPOQVYQ 2112
 Db 1682 --NNRLKILILEVYKTTAAVEETIGRHVILGILDRSSKSQSSASILIMSEAEVASKSVHE 1739
 QY 2113 DKTLPL-----PAOSSVGP-----AKAQPQTAQPSARQPOQPOQPSAQPE----- 2153
 Db 1740 EHTFVYDESLIPYSGSDMPNDINMMSKVTEEGTELSQRLVRSGFAGTEIDPNEBELMLN 1799
 QY 2154 VOTOPEVQOTVTSVSHVPSAOPTHAQSRSQVAAQS-----OPQSVQOQSVFVYQSPS 2208
 Db 1800 ISSRLQAAVAKLELAISSETSOLEHAKVYOTELMREFRKQKQATELSKQOEBELREHIE 1859
 QY 2209 QTRIRPSTPSQLSPOQ-----OSQVQTTT-----SQPIPIQPHTSLOTPSQ 2249

Db 1860 ESRAREQDLAVELSKAGVYIDGVADEKTLFERQIOEKTDIIDRLQELILCASNLQELIFAE 1919
Qy 2250 GQPOQPOVOSSTOTLSSQOTLQVSVSSPSRPQLOIQOPQVIAVPOLQOQVYLSQI 2309
Db 1920 -----OQOIOEBELLRS-----QKRAMKAAGP---VEQ-----QLOQTEKLMKE 1958
Qy 2310 QSQVNAQIOAQOSGVQOQIK-LQLPQIOQSSAVQTHQIOLN-----VYVOASVQEQLOK 2364
Db 1959 KLEVOQAKVRBDLOQKVALEIDVEEQVSRFIELOEKNTLMLOQNOALQEQLEK 2018
Qy 2365 VOQLRPOQOKKQOQOIEIKREHTLOASNOSEIOLQ-QVAMKANA-----VIEHLKQK-- 2416
Db 2019 MKRFLEQALIDREHEDVQOQEQLEQOLKVPFRQPISEHOTREVOQIAHLKERTDK 2078
Qy 2417 -SMTPAEENOMT-VCONVMKYILDKIDKEKQAKKRRRESEYQK-----2463
Db 2079 CSELLSKBOLQDIOERNEIEIKLEFVRLEQALLERDKHGAVAPKELSLQVOLA 2138
Qy 2464 -----RSQONATKLSALLFKHKEQURLAKRALLDK-DIQIEVO-----BELKR 2508
Db 2139 ERDAIDRKKEKETTINLEQOL-----EQFRELENKEVEVOQLHMOLEIQKKESTRLQLEQ 2194
Qy 2509 DLKIKKEKDLMOLOATVAAAPCPVTPVLPAPRPPSPRPBGVQHTGLSTPLPVA 2568
Db 2195 ENKLEFD-DMEKLGLA-----IKESDAMSTQDOHVL 2224
Qy 2569 SOK-----RRREBKD-----SSSKSKKKMIISTSKETKKTQKCYICKTPYDESKFYIG 2619
Db 2225 FGKFAQIIQEQVEIDQNLQWQYKLOQOLKITTDNKVIEKKNLIDLEI-----QIE 2277
Qy 2620 CDRCONWYGRQVGLIQSEBELIDEVPCQOSTEDAMTVLPLETKDYEGELKV--LRS 2677
Db 2278 C-----LMSQOECVK-----RRREIEIQLENEIVIKLOQELANIGOKTS 2316
Qy 2678 LOAKKAMPLEPVDNDAPDYGVYIKERMDLATMEERQVRQRYEKLTEFVANMTIIFDN 2737
Db 2317 MAHSHIS-----EADS-----LKHQDQVYIAKLALEQOQVEATANEEMTFMKNVLAKE 2363
Qy 2738 CRYNPSDSPFYQCAVLESFVQKLGKFRASKSHNNKLOS 2778
Db 2364 TNE-----KMNQLTQELFLSKRRESVEKIOS 2390
RESULT 14
US-09-724-676-68912
; Sequence 68912, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing.
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68912
; LENGTH: 3852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68912
Query Match 2.6%; Score 379; DB 5; Length 3852;
Best Local Similarity 17.2%; Pred. No. 0.016;
Matches 481; Conservative 477; Mismatches 993; Indels 850; Gaps 118;
Qy 305 DCVAVET-----QKNKP--YIRHEPIGYDRSRKKY-----FLNRLLIE 341
Db 113 DCSSEVNGCSFYWRTGKPTNLLREEFQVDSYSEQADSPHTLHMESELAGKHETE 172
Qy 342 EDTENENKKIYYSTKYVLAELIDCL-----DKDYWEAEICXII 381
Db 173 ELNRELEKRVYVYEGEGLOQLOFEFAIKQDGIITQULANIQARREDEMTREFLELT 232
Qy 382 EEMRE-ETI-RHMDITEDLTNRKARSGNSKSLAANEET-----LESIRAKKG 426

Db 233 EQSQKLOIQFOQLQASEETLKNSTHSSSTAADLOAKQOILTHQOQLEBDHLLDEYQKKE 292
Qy 427 DI-----DANKSPEETKDKNETENDSKDAEKNE-----EFEDQSLEKDS 467
Db 293 DFTMOISFLQEKIK-VYEMQDK-KVENSKEEIQEKETIIEELNKKIIEBKKTLEKD 350
Qy 468 DDKTPD-----DDEQEGSEVDFKSEKNGELSEPQAG--KQASGSTRITRLRNP 518
Db 351 KLTFAKRLGELQEQIYQKNQELIKNMKLELTNKKOKERQSESEIKQMGVELEQRNHK 410
Qy 519 DSL-----SOLSQOYAAAHAHA-NKLFKEGKVLVYNSOGELSRISTKKEVI 566
Db 411 DSOFEFDIVQRMQEQRKLEQRLAEIDEMYGQOYQMKQELLROHMAQMEKKTTRKGE 470
Qy 567 MKGNINNYFKLQEGKRYRVYHNOYSTNSFALNK--HOHREHDHQRHRLAHKFCUTPAGEF 624
Db 471 MENALNSYNI-----YNEDOIKLMMVAINELNKLQDTNSQKEKLEL-----516
Qy 625 KMGVSGSKVLITSLRLITQLENNISSFLHPNMAWASHRAMWIKRQVCMCRPREFALA 684
Db 517 ---GLIEEKCALQROLEDLEVEEL-----SFSREQ--IQARQITAEQD-SKLNEMKHS 564
Qy 685 LAILCEAVKPVMLPIWREFLGHTRLHRMTS-----IERBEKVKKKKKOEFEETMQOA 740
Db 565 LSTVEDLKAELIVASASESRKLELKHAEVYNYKIKLEMLEKERNAAVLDRMAESQEAELER 624
Qy 741 TWKTYTPRVKHWQKQGE---EYRVYGYGWMWIKTHYRRVPLRPGTNTNNYKSLLE 797
Db 625 LRTQLLFSHEELSKLEDELEIRIN-----IEKLMQIGHYKQIID 668
Qy 798 GTRKNNDEND--ESDKRKSRSRKKIKIEPDSEKDEVKSDAAKAD-----QNEM 847
Db 669 GLQNEKSQKLEWQFPRKDNITQKNQILIEISKLOQOSLVNSKSEEMTLOINELOKEI 728
Qy 848 DISKITEKK---DOVKEL-----LDSDDPKCKEEMPEY 879
Db 729 EILROEKEKGTLEQOELOLQTELEKQMKERENDLOEKFAOLBAE-NSILDKERTKL 787
Qy 880 DDMKTESHNOQESSQOVVYVNSSEGFHLRTSYKKTTSKSLDGLLEIRIKOTLEKOR 939
Db 788 EDMKLIHTPYSOE-----RLIFLDSIKSKSDSVWEKEL-ELLIEBEND 831
Qy 940 LEK--IKLEGGIKIGKGTGNSNKKNL-----SESPVITKAKECQSDSMRQE---Q 985
Db 832 LKQOQLOLNEIEIKQNTSPFAKNEFVANOQLEQETACLLKDDLEDSKKNQELEYS 891
Qy 986 SPANNDQEPDLIQCSQSDSVLRMSDPSHTTNKLYPKDRVLQVYSINSPEF-----1038
Db 892 KKLALNEE-----LHQRINPPTYVKM--KSSVDEDEKTFVAETLEMGEEV 934
Qy 1039 -----KCPKQNSTIENDIEEKVYSDLASGQEPYTSKGTGNPFIDDSK-LASADD 1086
Db 935 EKDQTELMKELEYTKREKLE--LSQRLSDIS---BQLQKQGEISFLVEEYVSLKPEKE 988
Qy 1087 ICTGLICKNNKPLIQEESDPTIVSSSKSALHSSVPKSTNDRDAPLBRAMPFEGKLGCDSES 1146
Db 989 PAVSLR-----ELEIITINHRA-----ENQOSDITQV 1016
Qy 1147 NSTLENSDPTVSIQDSSSEDMIVQNSNESTISEQFRTREDQVEYLEPLKCELVSGESTGNC 1206
Db 1017 SSL-----GGVYVMTSRGAGSVSKNKSFGESKIMVEDKVSFE---NMVGESEKQD 1068
Qy 1207 E---DRLP--VKGTEANGKKPQOQKLEBPVKKCDQIKLKATTKKNNENNESKQKQ 1261
Db 1069 QILIDHLPVYTVKSSLRATQPSENDKLQ-----KELNVLKSEQNDLR-LQMEQ 1116
Qy 1262 R-----TSFQI-----NGKDNKPKIYIKGECLKEISESVAGNPEPKNNINKIT 1308
Db 1117 RICLSLVYSTHVDQVREYENEND-----KALCSLKEBELIFA--QBEKKLEQK-I 1164
Qy 1309 PENDIKSLTVKSAIRPITNGD-----VIMEDFERNRSSEKTSKSHLLSSSDAEGNYSLS 1362


```

Db 1165 HOLELTQMTQET-----GDEGRPLHLIGLQKAVSECSFYLTQCLSGVEY---- 1213
QY 1363 ELPSTKESDSTQTTTPSPACSPESNSVNGVEDMEIETSEVKVTSSPITSEESNSLNDP 1422
Db 1214 -----YTPALKG-----EVNAEDKEN--SGDY 1233
QY 1423 IDENGLPINKN---ENVNGESKRRTVITEVT-----TMTSTVANEKSTVIAKVEKG 1469
Db 1234 ISENEDEPLQDYRYEYODFOENMHITLTKVTEENKLLVQTLRLSKIMGOOTQGMKLEFG 1293
QY 1470 DKQTVVSTENCASKSTVTTTTVTYTKLSTPSGSDIISVKQSKVTYTTTDSLTIT 1329
Db 1294 E-----ENLPKE-----ETEFJSHSQ----- 1310
QY 1530 GGLVTSMYTSKYSTRDVKYKLMKFSRPPKTRSGTALPSYRKFPVTKSKSIFLVPMDL 1589
Db 1311 -----MTNLEIDIVNHSKSLSDLEKTKLEBOVQELSLSSLOO-----L 1354
QY 1590 KTLARKGIREVVPYFNMAKRALDIMPSPRPFTGITWBYRLQTVKSLAGVSLMRLW 1649
Db 1355 KETEQ-----NYEAE-----IHCLO--KRLQAVS----- 1376
QY 1650 ASLRMDMAKAVPPGGSGTETSETETTELTKRRDVPYGRREYCIKTIICPIGPV 1709
Db 1377 -----ESTVP-----SLPDSVYLTESDAQRTMYPGS-----CYKKNI--DGTI 1414
QY 1710 ETPKETPTQPKRGLRSSALRPKRPETPKOTGPYIIETWVA--EEELMETRAFAERYEKE 1768
Db 1415 EFGESGVKEETIVKLEQYQOOLEEBAKIVSMSINFAQOTELSTRISGSKENASS 1474
QY 1769 K-AQAVEQAKRLEQOKPTVATSTTSPTSSTSTISPAQKVAVAFISGVTGTGMVL 1827
Db 1475 KOAHAAQOQOEHYFNEKK-----LSODQIGQFTEFYVDVYKFEKPLSKELGHEKEL 1529
QY 1828 TTVGSPATVTPQONNENHOFATWVKQGSNSGVQVQOKVGLIIPSSGTGQQTTSF 1887
Db 1530 LS-----NSDPHD-----IPESKUCVLT-----SEEMFSK 1556
QY 1888 QPRTATVITIRPNTSGSGGTSTNSQVITGPQIRPGMTVIRTPLOQSTLKAIIPTVPWCP 1947
Db 1557 KTFIVQSHIDEISVSMOASROLMEBLE-----DMQOE-----LVRO 1597
QY 1948 GADQVMTQILRGQVPYSTAVSADPNVSPFGOKSLTSANSTSNIOSSASGPPPPQOGVQ 2007
Db 1598 YQEHQOATELLR-----QAMHROMERQEDQ 1625
QY 2008 L--TMAQLT-QLTQGHG--NOGLTVIQQGGQTTQLOLIPQV--TVLPFGQQLMQAA 2061
Db 1626 LOEIRKRLNQLAQQNGENQOQ--EVEEQTFEREKELDRKPEDEVPEILSNERVALOKA- 1681
QY 2062 MPNGTVORFLEPLATTATTAFT-----TTTVSTTAAGTGBROSKLSPOMQVHQ 2112
Db 1682 --NNRLKILLEVYKTTAAVEETIGRHVIGLIDRSSKSQSSASLIMRSEEAIVKSCVHE 1739
QY 2113 DKTLF-----PAOSSVGP-----AKAQPOTAPSAHPQOTOPQSPAEP----- 2153
Db 1740 EHTRYADESIPYSQSDMPRNDIMNMSKVTEEGTELSQLRVSGFAGTEIDPENNELMN 1799
QY 2154 VQQPEVQVQTVYSHVPSAOPTHAQSCKPOVAAS--QPSVQOQSVYRQVQSPS 2208
Db 1800 ISSRLOAAVEKLELAISSTLSOLEHAKVYOTELMRSEFRQKQEAATESLQOEFLRLHE 1859
QY 2209 QTRIRSTESQLSPOQ-----OSOVQTTT-----SOPILQIPHTSLQIPSQ 2249
Db 1860 ESARARQALVELSKACGYIDGYADEKTLFERQIOEKTDIIDRLQELCLCSNKLQELAE 1919
QY 2250 GQPSQPOVQSSQTQLSSQOTLMQVSVSSPSRQLOIQOPQVIAVPLQOQOVYLSQI 2309
Db 1920 -----QQOIOEBRELLSR--OKEAMKAEAGP--VEQ-----QLQETEKLMKE 1958
QY 2310 QSQVVAQIQAGSGVPOQIK-LQPLQIQOQSSAVQTHQIONV-----VTVQAAVQOQOLR 2364
Db 1959 KLEVOQAEKVRDLOKQVKALEIDVEQVSRFIELEQEKNEIMDLRQONQALEKOLEK 2018

```

```

QY 2365 VQOLRQOQKKKQOQOIEIKREHTLOASMOSEIIOK-QVAKKHNA-----VIEHLKOKK-- 2416
Db 2019 MRKFLQEQALDRBHERNVPQEOIQKLEQOLKVPYPRQPISEHQOTREVEQALNLIKETDK 2078
QY 2417 -SMTPAERENQMT-VQNOVMKYIILDKIDKEKQOAKKRRSESVQK----- 2463
Db 2079 CSEILLKSEQLOIDQERNEIEKLEFRVRELQALLQEDKHHGAAVEAKPSLEVOLOA 2138
QY 2464 -----RSQONATKLSALFLKHKQOALFKKRALIDK--DLQIEVO-----BELKR 2508
Db 2139 ERDAIDKREKEITNLEBOL-----EQPRELEKNNEVOQLHMOLEIQKKESTRLOBLEQ 2194
QY 2509 DLKIKREKIDLMQLOATAVAAPCPVTPVLPAPAPPPSPPPGVOHTGILSTPLPVA 2568
Db 2195 ENKLFND-DMEKIGLA-----IKESDAMSTQDQHYL 2224
QY 2569 SQK-----KRREBKD-----SSSKSKKKMMISTTSKETTKKDTLYCICKTPYESKFYIG 2619
Db 2225 FGKFAQIIOKEVEYIDQLNBOVTKLOQOLKITTDNKKVIEKKNELIRDLFT-----QIE 2277
QY 2620 CDRQOMNYHRCVGILOSEAEILDEYVCPQCSQSTEDAMTVLFTLTKDYEGLKRV--LRS 2677
Db 2278 C-----LMSDOQVK-----RNREBELQNEVIEKLOQELANIGOKTS 2316
QY 2678 LOAHKMAPLEVPDNDADYGVIVKEPMDLATMERVQRRYVEKLTREVADMTKIFDN 2737
Db 2317 MNHNSLS-----EADS-----LKHQLDVYIAELALQEQVETANEMFTMKVNLKE 2363
QY 2738 CRYNPSDSFPYQCAEVLESFPYQKLGKFGKASSHNNKLOS 2778
Db 2364 TNF-----KMNQLOTEFLSKRRESEVETIOS 2390

```

```

RESULT 15
US-09-724-676-68913
; Sequence 68913, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68913
; LENGTH: 3852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68913

```

```

Query Match          2.6%; Score 379; DB 5; Length 3852;
Best Local Similarity 17.2%; Pred. No. 0.018;
Matches 481; Conservativity 477; Mismatches 993; Indels 850; Gaps 118;

```

```

QY 305 DCVAEI-----QKNRP--YIRHEPIGYDRSRKYV-----FLNRLLIE 341
Db 113 DCSSEVAGCSFVMRTGPTLLRLREERGVDDSYSEGAQDSPTHLMESELAGKHQHEI 172
QY 342 EDTENENEKKIWTYSTRKVLAEILDCI-----DKDYEAELCKIL 381
Db 173 ELNRELEEMRVATYGTGLOQLOEFAIKORDGIITQLTANLQOARREKDETRRELELT 232
QY 382 EEMRE-EIH-RHMDITEDLLNKARGSKSPFLAANEI-----LESIRAKG 426
Db 233 EQSQKLOIQFOQLASSTLNSHSTAAADLQAKQOILTHQOOLEBODHLLEDYQKKE 292
QY 427 DI-----DNVKSPEETEKDNENETENSKDAEKNE-----EFEDQSLKDS 467
Db 293 DFTMQISFLQEKIK-VYEMQDK-KVENSNSKEILOKETTIEELNKKIIEKKTKLEMD 350
QY 468 DDKTPD-----DDPEQKSEVGVDFKSEKSNGLSESPQAG--KASGSTRITIRLRND 518

```


Db 351 KLTADKLLGLOEQIVQNOEIKNNKLELTNSKÖKEÖSSEIKOLMGTVELOÖRNNK 410
 QY 519 DSKL-----SÖKSÖQVAAAHAHA-NKLFEGREVLVNSOGELSRSTKEVI 566
 Db 411 DSOEFDIYORMEÖEQÖRLQELAEDEWQÖIVQMKOELLRÖMAÖMEKTRHKE 470
 QY 567 MGNINNYKLOEGKYRYHNOYSTNSFALN--HÖHEDHDKRHNAHFCFLTPAGEF 624
 Db 471 MEWAALSYNI-----TYNEDOIKLMAVAINELNIKLODTNSÖKEKLEEL-----516
 QY 625 KNGSVHSGSVLITSLRLITQLENNIPSPFLHPMAWSHRAWIKAYOMCSKPREPALA 684
 Db 517 ----GLIEEKALÖRÖLELVEEL-----SFSREQ--IÖRÄQTÖAEÖE-SKLINEAHKS 564
 QY 685 LALIECAVAPVYMLPIWREFLGHTLRMTS----IEREKEKVKKREKÖEETMOA 740
 Db 565 LSTVEDLKAIEIVASBSREKLELHEAVTNKIKLEMEKEKNALVDRABESÖEALER 624
 QY 741 TWVKTFEPYKHQWOKGE---EYRTYGGGMSWISKTIVYRFVPRLPQNTNVNYSLE 797
 Db 625 LRTÖLFSHEEELSKLEDELEIHRIN-----IEKLDNLTGIHYKÖOID 668
 QY 798 GTRNNNDENMD--ESDKRCRSRPPKIKIEPSEKDEYKSDAKGAD-----ÖNEM 847
 Db 669 GLÖNEMSÖKIEFMÖPEKDNLITKÖNÖLIEISKLDLÖÖSVNSKSEEMTLOJNELÖKEI 728
 QY 848 DISKITEKK---DÖDVKEL-----LDSDSKPCKEPEMEV 879
 Db 729 ELTROBEKEKGTÖEÖVÖLÖKTELLEKÖMEKENDLÖEKFQÖLEAE-NSILKDEKTL 787
 QY 880 DDDMKTESIVNÖESSÖUVNVANVSEGFHLRTSYKKKTKSKLDGLERIKÖFTLEKÖR 939
 Db 788 EMLMLHPVSÖE-----RLIFLDSIKSKSDSWEKEI-ELLIEENED 831
 QY 940 LBEK--IKLEGJIKGJIKTSTNSKNL-----SESPYITAKBECCÖDSMOE----Q 985
 Db 832 LKÖCÖGLÖNEBELKÖRNTSPFAKNEFVYQÖLEÖEYACLVKÖDLEDSKNÖÖLEYKS 891
 QY 986 SPANANDÖPÖELLÖCÖSÖSSVLRMSDPSHTNKLPRKDRYLDVYSIRSPEI-----1038
 Db 892 KKLALNEE-----LHÖRINPTTVKM--KSSVFDEKKTVAELTEMGVAV 934
 QY 1039 -----KCPKÖNSTIENDIEEKVSDLASRGÖPTSKTGNDFTIDSK-LASADD 1086
 Db 935 EKDTTELMKEKLEVTREKLE--LSÖRLSDLS---EÖLKÖKPELISFLINEEYKSLPEKE 988
 QY 1087 IGTLLCKNNKPLIEESDPTVSSKSALHSVPKSTNDRDAPLSRAMFEGKLGDSKS 1146
 Db 989 PYSLCKR-----ELEIINHNA-----ENVOÖCDTÖV 1016
 QY 1147 NSTLENSDPTVÖSÖSSEEDMTIVÖNSNESISÖFRTREÖDVEVLELCELVSGESTGNC 1206
 Db 1017 SSSL-----DGVYTMTRSÖRGEÖSVKNSFGESKIMVDKVSFE--NMIVGEÖSKÖE 1068
 QY 1207 E---DRLP--VAGTEANGKPKSÖQKKLEBPVKNCSÖDÖIKLKNNTDKKNENRESEKÖQ 1261
 Db 1069 ÖLILDLPSVTRESSLRATÖPÖENDKLÖ-----KELNVLKSEÖNDR-LÖMEAO 1116
 QY 1262 R-----TSTFOI-----NGCKNKPKYTLKGECLKEISESRVYSCNPEPVNINKI 1308
 Db 1117 RCLSLVSTHVDÖYREVMENKÖD-----KALCSLKEELIFRA--ÖEEIKELÖK-I 1164
 QY 1309 PENDIKSLVKEASIRPITNGD-----VIMEDFNERNSSETFKSHLSSSDAGNÖRDL 1362
 Db 1165 HÖLEIÖTQMKTÖE-----GDEGRPLHLLIÖKLOKAVSECSYFLOTLCSVLGEI-----1213
 QY 1363 ETLPSKESDSÖTÖTTPASCPESNSVNOVEDMEIETSEVKVNTSSPITSEESNLNDP 1422
 Db 1214 -----YTPALKC-----EYVAEDKEN-SGDY 1233
 QY 1423 IDENGLPINKN---ENVNGESKRTVITEVT-----TMSTVATESKVIKVEG 1469
 Db 1234 ISENDEPÖLÖDRYEVÖDÖENMHÖTLNKNVTEENKLLVLOTRLSKIMWÖQÖTÖGKMLEG 1293

QY 1470 DKÖTVVSTENCAKSTVTTTTVTRKLSPTSGSVNDISVKÖÖSKTVTTTTVTDLSITTT 1529
 Db 1294 E-----ENLPE-----LEPELISHÖ-----1310
 QY 1530 GGLTVLTMVYSKEYSTRDKYKLMKFSRPPKTRSGILPSTYKRVYTSKSIPLPNDI 1589
 Db 1311 -----MNLBÖIDVNHHSKSLSDLEKTKLEBÖEÖLESILSSLOÖ-----L 1354
 QY 1590 KKLARKGIREVYEVYNNKPALDIMPSPRPTPGITMWRRLÖTVKSLAGVSLMLRLM 1649
 Db 1355 KTEÖ-----NYEA-----IHCLÖ--KRLÖAVS-----1376
 QY 1650 ASLRMDMAKVPYGGGSGSTRETSETETTELIRKRDVGYIREYCIKRICIGVP 1709
 Db 1377 -----ESTVPP-----SLPVDVSVITTESDARTMPPS-----CYKKNI--DGTI 1414
 QY 1710 ETPKEPTPÖRKGLRSSALPRKRETPKÖTGYIETWA--EELELMEIRAPAEVEKE 1768
 Db 1415 EFSGEFVKEETNIVKLEKÖYÖÖLEEVAKYVMSIAFAÖÖTBLSRISGGKENTASS 1474
 QY 1769 K-AÖAVEÖÖAKRRLBÖÖKPTVATSTPSSTSTISPAÖKVMAPISGSVTTGKML 1827
 Db 1475 KÖAHAVCÖÖBÖHYFENMK-----LSÖDÖIGFÖTFEVÖVYKFEKFRPLSELGEHÖKEIL 1529
 QY 1828 TTKVSPATVTFÖÖKNFHTÖFATWVKÖGÖSNGVÖVÖÖKVGLIPSSGTSGÖÖFTSF 1887
 Db 1530 LS-----NSDPHD-----IPESNDVILTI-----SEMFNSD 1556
 QY 1888 ÖPRTATVTRPMTSGSGGTTNSÖVITGPÖIRPGMTVTRPLOÖSTLKAIRTPVWOP 1947
 Db 1557 KTFIVRÖSHIDEISVSMASRÖLMNEÖLE-----DMRÖ-----LVKÖ 1597
 QY 1948 GAPÖVYMQÖIRGÖPSTAVSAPNTVSSTRPÖKSLTSATSTNSIGÖSASÖPPRÖÖGÖVK 2007
 Db 1598 YÖEHÖÖATELRL-----ÖAHMRÖMEKÖRÖDÖÖ 1625
 QY 2008 L--TYAÖLT-ÖLTOGHG--NÖGLTVVIOGÖGÖTÖLÖLIPÖV--TVPRGÖQÖLÖMÖA 2061
 Db 1626 LÖEIKRLRÖLÖAÖÖGNENÖG---EVEÖTPEKELDKKPEVPRILLSENÖYALÖKA- 1681
 QY 2062 MPNGVÖRPLFTPLATTATTAFTAST-----TTTTVSTAAÖTGÖRÖSKUSPÖMÖVHÖ 2112
 Db 1682 --NNRLKILLEVKTAAVEETIGRHVGLIDRSSKÖSÖSASLIMRSAEASVKSÖVHE 1739
 QY 2113 DKTLR-----PÖÖSSVGV-----AKAÖPÖTÖASAPÖQÖTÖQÖSAPÖE-----2153
 Db 1740 EHTRYTDESIPSTSGDMPRNDINMWSKVTERGETELSQRLVNSGFAGTEIDPNEELMUN 1799
 QY 2154 VÖTÖPEVÖTÖTVVSHVPEAÖPTHAÖSSKPOVAAÖS-----ÖPÖSNGÖSPVRVQÖSPS 2208
 Db 1800 ISSRLÖAÖVKEKLELAEISÖSSÖLEHAKYVÖTÖLMRSPFÖKÖEATPSLSCÖBÖELRRLHE 1859
 QY 2209 ÖTRIRPSPÖLSPO-----ÖSÖVÖTÖT-----SÖPIPTÖPHTSLÖIPSO 2249
 Db 1860 ESARÖLÖLAVELSKAEGVIDYADEKTELEFEROIÖEÖTIDILRÖELLCASRLÖELAE 1919
 QY 2250 GÖPÖSÖPÖVÖSSÖTÖLSSÖQÖTLÖNÖVSÖSPSRPÖLÖIÖQÖPÖVYLANVPLÖÖQÖVÖVLSÖI 2309
 Db 1920 -----ÖÖÖIÖERELISR---ÖKEAMKAEAP--VEÖ-----ÖLÖÖTEKELME 1958
 QY 2310 ÖSÖVVAÖIÖAÖÖSGVÖPÖIK-LÖLPÖIÖÖSSAVÖTÖHÖIÖV-----VYÖAÖSÖBÖLÖR 2364
 Db 1959 KLEVÖCÖAKVNDLÖKÖYKALAEIVÖEVSÖFIEÖKNTÖELMDLÖRÖNÖALÖKÖLEK 2018
 QY 2365 VÖÖLRÖÖÖKÖKÖÖÖLEIKREHTLÖASNÖSEIÖK--ÖVVMKNA-----VIEHLÖKÖK-- 2416
 Db 2019 MKRFLDEÖALDIREHNERDVÖÖEÖIÖKÖLÖQÖLVPRPÖISEHÖTREVÖÖLANHLEKÖTDK 2078
 QY 2417 -SMTAERENÖRM--VCÖVÖKVIÖLÖDKÖEÖKÖAKKREÖEVSÖK-----2463
 Db 2079 CSÖLLLSKÖLÖRDÖÖERNEIEKLEFRVRLÖEÖALLÖBRKHFGAÖVAPÖELSLÖVÖLÖA 2138

